Inflammatory Diseases – An Analysis of the Underlying Biological Triggers

Bennike, Tue Bjerg

DOI (link to publication from Publisher):
10.5278/vbn.phd.engsci.00012

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

PROTEOME RESEARCH

INFLAMMATORY DISEASES – AN ANALYSIS OF THE UNDERLYING BIOLOGICAL TRIGGERS

BY
TUE BJERG BENNIKE

DISSERTATION SUBMITTED 2014

AALBORG UNIVERSITY
DENMARK

INFLAMMATORY DISEASES – AN ANALYSIS OF THE UNDERLYING BIOLOGICAL TRIGGERS

BY
TUE BJERG BENNIKE

AALBORG UNIVERSITY
DENMARK

DISSERTATION SUBMITTED 2014

Thesis submitted: 2014

PhD supervisor: Allan Stensballe
Aalborg University

PhD committee: Professor Steffen B. Petersen (Chairman)
Aalborg University, Denmark

Professor, Bent Honoré
Aarhus University, Denmark

Professor Luca Bini
Universita di Siena, Italy

PhD Series: Faculty of Engineering and Science, Aalborg University

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-413-2

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2015

Thesis title:

Inflammatory Diseases – An Analysis of the Underlying Biological Triggers

PhD student:

Tue Bjerg Bennike

Supervisors:

Allan Stensballe, PhD, Associate Professor

Svend Birkelund, MD, PhD, DMSc, Professor

Vibeke Andersen, MD, PhD, Associate Professor

List of papers included in the PhD thesis:

[1] **Bennike, T.**, Birkelund, S., Stensballe, A., and Andersen, V. Biomarkers in inflammatory bowel diseases: Current status and proteomics identification strategies. World J Gastroenterol 20, 3231–3244 (2014).

[2] **Bennike, T.**, Lauridsen, K. B., Meyer, M. K., Andersen, V., Birkelund, S., and Stensballe, A. Optimizing the Identification of Citrullinated Peptides by Mass Spectrometry: Utilizing the Inability of Trypsin to Cleave after Citrullinated Amino Acids. Journal of Proteomics & Bioinformatics 06, (2013).

[3] **Bennike, T.**, Ayturk, U. M., Haslauer, C., Froehlich, J. W., Proffen, B. L., Barnaby, O., Birkelund, S., Murray, M. M., Warman, M. L., Stensballe, A., and, Steen, H. A normative study of the synovial fluid proteome from healthy porcine knee joints. Journal of Proteome Research (Accepted, August 27th 2014).

NB: Included papers are bold and underlined in the text.

This thesis has been submitted for assessment in partial fulfillment of the PhD degree. The thesis is based on the published scientific papers which are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and are also available at the Faculty. The thesis is not in its present form acceptable for open publication but only in limited and closed circulation as copyright may not be ensured.

Contents

Preface

The present PhD thesis is based on clinical research I have performed as a PhD fellow in the period September 2011 to August 2014. The work was carried out at the Laboratory for Medical Mass Spectrometry, Department of Health Science and Technology and Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, Denmark, and at the Steen and Steen Laboratory, Proteomics core facility, Boston Children's Hospital, Harvard Medical School, USA.

I would like to thank Associate professor Allan Stensballe, Professor Svend Birkelund, and Associate professor Vibeke Andersen for giving me the opportunity to work on this great project and providing steady and reliable supervision. A special thanks to Professor Gunna Christiansen for providing additional and essential supervision. Also, I would like to thank Associate professor Hanno Steen for accepting me in his laboratory and group at Harvard Medical School.

I would like to thank my colleagues involved in the main project for a great ongoing collaboration, including: Thomas Gelsing Carlsen, Kasper B. Lauridsen, Henning Glerup, Torkell Juulsgaard Ellingsen, Ole Kristian Bonderup, Lau Sennels, and Martin Bøgsted.

Additionally, I would like to thank all of my colleagues, past and present, at Aalborg University and at Steen and Steen Laboratory, including: Michael Kruse Meyer, Henrik Kjeldal, Florian-Alexander Herbst, Rasmus Hansen, Anne Louise Schacht Revenfeld, Charlotte Sten, Anne Sommer, Ditte Bech Kristensen, Kevin Broadbelt, Jan Müntel, Waltraud Mair, Omar Barnaby, John Froehlich, Ugur Ayturk, Carla M. Haslauer, Saima Ahmed, John Sauld, and Ali Ghoulidi.

I would like to thank my family and friends for their support and encouragement. Last but not least, thank you Kathrine Bjerg Bennike for supporting me – and for being such a lovely wife.

I hereby declare this is my original work.

_____

Tue Bjerg Bennike

Abstract

The etiology of the inflammatory diseases, and what initially triggers the immune system to target otherwise healthy cells and tissue, remains poorly understood. The undesirable immune response often results in chronic conditions, leading to life-threatening complications for the affected individuals, as well as a significant economic burden for society. This PhD thesis aims at gaining a deeper understanding of the underlying biological triggers of the autoimmune diseases by using the inflammatory joint disease rheumatoid arthritis and the inflammatory bowel disease ulcerative colitis as model systems. The two diseases share several genetic factors and environmental risk factors, suggesting common triggers. The posttranslational modification citrullination is known to be involved in the etiology of rheumatoid arthritis, and might be involved in inflammatory bowel disease as well. However, the identification of the modification is often non-trivial.

The main analytical method used in this PhD was mass spectrometry driven protein identification and relative quantification. In order to enable the reliable identification of citrullinated peptides and proteins, I analyzed several synthetic citrullinated peptides and synovial fluid from a rheumatoid arthritis patient. Protocols for the collection of mucosal colonic biopsies, serum, and plasma from 10 ulcerative colitis patients, 11 rheumatoid arthritis patients and 10 controls were established, and the biological samples from patients and controls were collected in collaboration with Silkeborg Regional Hospital. Protocols for the sample preparation and analysis of the colonic biopsies were optimized, and the samples were analyzed using high-throughput bottom-up proteomics.

The histological analysis revealed no visible differences between the colonic biopsies from the rheumatoid arthritis patients and the colonic biopsies from our controls. The crypt architecture of the colonic biopsies from the ulcerative colitis patients was preserved. However, the colonic tissue was found to have a higher density of cell nuclei, likely neutrophils, which was confirmed by the proteomics analysis. Of the 5,444 quantifiable proteins, 46 proteins demonstrated a statistically significant abundance change between the ulcerative colitis group and the controls, of which 33 proteins were more abundant and 13 proteins less abundant. Several of the statistically significantly more abundant proteins were involved in the innate immune system, and known components of neutrophil extracellular traps. Both findings are in agreement with the ulcerative colitis diagnosis, and indicates that even though colonic tissue from well treated patients was analyzed, an

inflammatory condition is present in the tissue. The inflammatory condition and activated innate immune system, indicates the involvement of the gut microbiota in the disease.

No proteins demonstrated a statistical significant abundance change between the rheumatoid arthritis group and the control group in the present study. However, the analysis of citrullinated peptides and proteins allowed us to propose an identification strategy for reliable identification of citrullination. Unambiguously, citrullination of arginine was found to result in a tryptic missed cleavage. Combined with the accurate peptide mass increase and peptide fragment masses, a reliable identification could be obtained. Using the identification strategy, one citrullinated peptide, the C-terminal end of the protein vimentin was identified with a statistical significant abundance change between the colonic biopsies from the rheumatoid arthritis patients and the controls. No citrullinated peptides demonstrate a statistical significant abundance increase between the ulcerative colitis group and the control group.

The identification of increased abundances of citrullinated vimentin in the colonic tissue of rheumatoid arthritis patients underlines the role of citrullination and vimentin in the etiology of rheumatoid arthritis. Citrullinated vimentin has previously been found in increased amounts in the lung-tissue of rheumatoid arthritis patients, suggesting this to be the initial place of autoantigen presentation in the disease. However, the finding of increased amounts of the same citrullinated protein in the gut could indicate that initial disease onset can take place in many places. Other studies have found indications of the gut microbiota being involved in the etiology of rheumatoid arthritis. No citrullinated peptides were found demonstrating a statistically significant abundance change between the ulcerative colitis patients and controls. This study thereby did not find indications that citrullination is involved in the etiology of ulcerative colitis. However, the activated innate immune system in the colonic tissue of ulcerative colitis patients and the likely increased abundance of neutrophil extracellular traps strongly indicate that the gut microbiota is involved in the etiology of ulcerative colitis. A study incorporating metabolomics and transcriptomics data would allow us to investigate the ongoing host-microbial interaction and altered gut microbiota functions. Such data is likely to provide new insights in the etiology of the inflammatory diseases, and potentially lead to new treatment strategies and diagnostic tools.

Dansk Resumé (Abstract in Danish)

De ætiologiske faktorer bag inflammatoriske sygdomme, og hvad der forsager immunsystemets angreb på rask væv og celler, er forsat et område, hvor der mangler en dybere forståelse. Det uhensigtsmæssige immunrespons resulterer ofte i kroniske tilstande, der fører til livstruende komplikationer for de berørte personer, samt en betydelig økonomisk byrde for samfundet. I denne PhD-afhandling forsøger jeg at opnå en dybere forståelse af de udløsende biologiske faktorer for de autoimmune sygdomme ved at anvende den inflammatoriske ledsygdom reumatoid artritis og den inflammatoriske tarmsygdom colitis ulcerosa som modelsystemer. De to sygdomme har adskillelige genetiske og miljømæssige risikofaktorer til fælles, hvilket indikerer at der kunne være fælles biologiske bagvedliggende faktorer. Motiveret af den kendte sammenhæng mellem reumatoid artritis og proteiner med den posttranslationelle modifikation citrullinering, undersøges det om citrullinering kan have en forbindelse med inflammatoriske tarmsygdomme.

Den primære analysemetode har i dette PhD-projekt været massespektrometri-baseret protein identifikation og relativ kvantificering (proteomics). For pålideligt at kunne identificere citrullinerede peptider og proteiner, analyserede jeg adskillelige syntetisk citrullinerede peptider og ledvæske fra en reumatoid artrit patient. Mucosale biopsier fra colon, serum samt plasma blev indsamlet fra i alt 10 patienter med colitis ulcerosa, 11 patienter med reumatoid artrit og 10 kontroller i tæt samarbejde med Silkeborg Regional Hospital, og biopsierne blev analyseret med high-thrughput bottom-up proteomics, ud fra optimerede protokoller.

De mucosale colonbiopsier fra reumatoid artrit patienterne var histologisk identisk med biopsierne fra kontrollerne. Dette var ikke tilfældet med biopsierne fra colitis ulcerosa patienterne, som generelt havde en højere tæthed af cellekerner. Dog var krypt arkitekturen velbevaret, hvilket stemmer overens med at kun velmedicinerede patienter var inkluderet i studiet. Den øgede cellekerner-tæthed er sandsynligvis neutrofile celler, hvilket blev bekræftet af proteomics analysen som er i overensstemmelse med patienternes diagnose. Af de 5.444 kvantificerbare proteiner, var der statistisk signifikant forhøjede mængder af 33 af proteiner i colitis ulcerosa-biopsierne i forhold til kontrolbiopsierne. Derudover var der statistisk signifikant mindre af 13 proteiner. Proteinerne med øgede mængder var forbundet med det medfødte immunsystem, og adskillelige af dem var desuden en del af de neutrofile ekstracellulære traps. Studiet indikerer at selvom colon-biopsierne stammede fra velbehandlede patienter er vævet i en inflammatorisk tilstand, og resultaterne peger på at tarmens mikrobiota i colitis ulcerosa er medvirkende til denne tilstand.

I dette studie udviste ingen af proteinerne en statistisk signifikant mængde-forskel mellem reumatoid artrit biopsierne og kontrolbiopsierne. Studiet af de citrullinerede peptider og proteiner viste at et citrullineret arginin uden undtagelse resulterer i et missed cleavage når proteinfordøjelsen foregår med trypsin. Det tryptiske missed cleavage kan med den korrekte masseforøgelse og peptid fragmentmasser bruges som en specifik markør for citrullinering, og baseret på resultaterne har vi foreslået en strategi for pålidelig identifikation af citrullinering. Ved at anvende identifikations strategien blev ét citrullineret peptid fundet med statistisk signifikant øgede mængder i mucosal biopsierne fra reumatoid artrit patienterne i forhold til kontrolbiopsierne. Peptidet tilhørte den C-terminale ende af proteinet vimentin. Ingen citrullinerede peptider blev fundet med statistisk signifikant forskel mellem colitis ulcerosa biopsierne og kontrolbiopsierne.

De forhøjede mængder citrullineret vimentin i colonvæv fra patienter med reumatoid artritis, understreger citrullinerings og vimentins rolle i reumatoid artrit. Et andet studie har påvist forøgede mængder C-terminalt citrullineret vimentin i lungerne på reumatoid artrit patienter, hvilket indikerer at tilstanden kunne have sin begyndelse i lungerne. De forhøjede mængder citrullineret vimentin i mucosal biopsierne fra colon peger imidlertid på, at sygdommen kan have sit udspring flere steder i kroppen. Ingen citrullinerede peptider blev fundet med statistisk signifikant ændrede mængder mellem colitis ulcerosa patienter og kontroller, hvilket indikerer at citrullination ikke er involveret i colitis ulcerosa. Resultatet af den øgede mængde af adskillelige proteiner forbundet med det medfødte immunsystem samt tegn på forøgede mængder neutrofile ekstracellulære traps i colonvæv tyder på, at tarmens mikrobiota er involveret i colitis ulcerosa. Et udvidet studie der inkluderer metabolsk data og transcriptionel data, vil kunne undersøger interaktionerne mellem immunsystemet og tarmens mikrobiota yderligere, og herigennem klargøre mikrobiotaens rolle i sundhed og sygdom. Denne viden kan potentielt føre til nye behandlingsstrategier og diagnostiske værktøjer for de inflammatoriske sygdomme.

Abbreviations

2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
A280	Absorbance at 280 nm
ACN	Acetonitrile
ACPA	Anti-citrullinated peptide/protein antibodies
AGC	Automatic gain control
ANA	Antinuclear antibodies
Anti-CCP	Anti-cyclic citrullinated peptide antibody
ASCA	Anti-saccharomyces cerevisiae antibody
BCA	Bicinchoninic acid assay
BSA	Bovine serum albumin
CD	Crohn's disease
CID	Collision-induced dissociation
CRP	C-reactive protein
DDA	Data-dependent mode of acquisition
ESI	Electrospray ionization
ETD	Electron-transfer dissociation
FA	Formic acid
FASP	Filter aided sample preparation
FDR	False discovery rate
FFPE	Formalin fixed paraffin embedded
H&E	Hematoxylin and eosin
HCD	Higher-energy collisional dissociation
IBD	Inflammatory bowel disease
IL	Interleukin
LC	Liquid chromatography
m/z	Mass-to-charge ratio
MS	Mass spectrometry
NET	Neutrophil extracellular trap
pANCA	Perinuclear anti-neutrophil cytoplasmic antibodies
PAD	Peptidylarginine deiminase
PCA	Principal component analysis
PMN	Polymorphonuclear leukocytes
PTM	Posttranslational modification
PSM	Peptide spectral match
RA	Rheumatoid arthritis
SDS	Sodium dodecyl sulfate
SWATH	Sequential window acquisition of all theoretical mass spectra
TCEP	Tris(2-carboxyethyl)phosphine
UC	Ulcerative colitis
UPLC	Ultra high performance liquid chromatography

1. Introduction

The immune system plays a key role in protecting our body from internal and external threats. By recognizing and neutralizing viral and prokaryotic pathogens and abnormal cells, including cancerous, we are to a large extent protected from infections and cancers. However, when the immune system targets otherwise healthy cells and tissues and produce autoantibodies, it leads to several types of autoimmune diseases. This undesirable immune response can result in inflammation and damage to tissues and bone, and may be specific to a particular type of cells, e.g. pancreatic β cells in type-1 diabetes and oligodendrocytes in multiple sclerosis[4]. A Danish study of 31 autoimmune diseases in 2007, found a prevalence of 5%, and the incidence and prevalence are increasing in developed countries[5,6]. The triggers of the initial immune response is not always known, and the etiology of the autoimmune diseases is incomplete. However, genetic factors and environmental stimuli, such as smoking, has been found to predispose to the development of several autoimmune diseases[4].

This PhD thesis aims at gaining a deeper understanding of the underlying biological triggers of the autoimmune diseases, using inflammatory joint and intestine diseases as a model system; namely inflammatory bowel disease (IBD) and rheumatoid arthritis (RA). IBD and RA disorders can lead to life-threatening complications, such as stroke events and colon cancer for RA and IBD, respectively. Furthermore, the diseases have a major impact on the quality of life for the affected individuals, and for society due to lost labor and expenses to the health care system[7–9]. Several genetic risk loci are shared between IBD and RA[4], and the two diseases share many similarities with other autoimmune diseases[10]. Several common inflammatory inducing cytokines are known to be involved in the disease etiologies of IBD and RA, including T-helper 17 cells cytokines interleukin (IL)17 and IL23[11,12]. Smoking is a risk factor for Crohn's disease (CD), one of the two most common forms of IBD, and for RA where it is the environmental factor with the highest effect on risk[13,14]. The cause is unknown; however, increased expression of the enzymes peptidylarginine deiminases (PADs) have been found in the lungs of smokers[15]. PADs catalyzes citrullination, a posttranslational modification (PTM) of proteins, and also increased amounts of citrullinated proteins have been found in the lung-tissue of smokers[15]. It has been hypothesized that the citrullinated proteins may become an autoantigen and thereby trigger an immune system response in people with genetic dispositions for RA and IBD[16–18].

1.1. Inflammatory Bowel Disease

IBD comprises multiple chronic gastrointestinal disorders, and the two most common forms are CD and ulcerative colitis (UC). CD is characterized by chronic inflammation in any part of the gastrointestinal tract from mouth to anus. Most commonly, the terminal ileum or the perianal region are inflamed and in a non-continuous manner (Figure 1). Histologically, CD shows thickened submucosa, transmural inflammation, fissuring ulceration and non-caseating granulomas. UC is characterized by inflammation limited to the colon, spreading continuously from the rectum and various distance proximal. Histologically, UC shows superficial inflammatory changes limited to the mucosa and submucosa with inflammation of crypts (cryptitis) and crypt abscesses[19]. The etiology of IBD remains unclear, but genetic studies have found 163 genetic loci involved in IBD, many of which pointing to defects in the hosts defense against infections[20]. Based on this and other studies, IBD likely involves a complex interplay between genetic and environmental factors, including the gut microbiota[21,19,22–25,20].

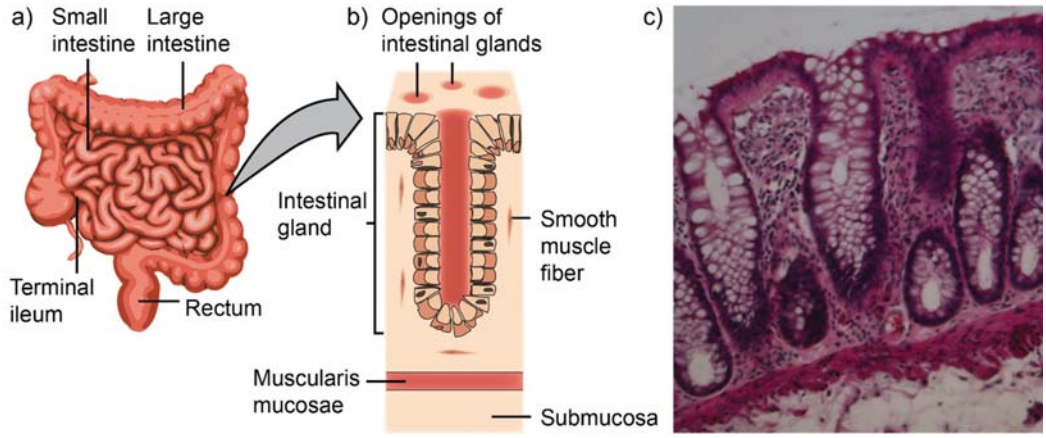


Figure 1. Overview of the human intestine. a) Small and large intestine. b) Cross section of human colon (large intestine) with epithelial cells. c) Hematoxylin and eosin (H&E) stained large intestine biopsy from the present study.

The incidence and prevalence rates of IBD vary considerably between countries, with higher occurrence in industrialized countries than in non-industrialized countries. The highest occurrences are found in North America and Europe, and within Europe the highest occurrences are found in Scandinavia and the United Kingdom[26]. It is estimated that 0.3%-0.8% of the northern European population suffers from IBD[27–29]. Furthermore, the prevalence of IBD is higher in females and increases with age[30]. As of 2013, the incidence rates of both UC and CD are stable or increasing in most countries, thus the number of IBD patients is strongly expected to increase in the future[31]. IBD greatly reduce the quality of life of the affected individuals, and the direct annual expenses to the European health care system is as of 2013 estimated to be 4.6-5.6 billion EUR[26].

1.1.1. Diagnosis of Inflammatory Bowel Disease

Several general markers for inflammation obtainable in stools or blood samples are used in the diagnosis of IBD (Table 1), and the diagnosis is often verified by endoscopy followed by histology. However, as outlined in our recent review no biomarkers capable of distinguish CD from UC patients with a high specificity and sensitivity have been implemented in daily clinical use[1,32–35]. Instead, the diagnosis is established based on a combination of symptoms, clinical examinations, laboratory findings, radiology, and endoscopy with histology, which are used to assess disease severity. Even when the tests are performed by expert clinicians, they can for some patients result in diagnostic uncertainty[34–37]. Hence, there is a need for reliable and usable biomarkers for early diagnosis and prognosis of IBD, which could be based on disease specific proteins[1,22,32,35,37–40].

As the intestine is the site of inflammation, the intestinal tissue remains an obvious place to search for novel disease specific proteome alternations for IBD, usable for explaining disease etiology. Subsequently, easily obtainable biological material such as serum, plasma, feces, and urine can be screened for the identified disease specific markers to assess the usability in blood screenings as novel diagnostic markers. Several proteomic studies have been conducted with the aim of identifying such disease specific markers using proteomics. However, as we stated highly promising biomarker candidates have yet to be identified and implemented[1].

Biomarker	Specificity	Usability
Serum Biomarkers		
ASCA	39-79% of CD patients positive, 5- 15% UC patients[41–43].	14-18% of controls tested positive, limiting the diagnostic value[44].
pANCA	20-85% of UC patients positive, 2-28% of the CD patients[41,43,45].	32% of controls tested positive, limiting the diagnostic value[44].
CRP	Marker for acute inflammation.	Cannot differentiate CD from UC. However, usable for monitoring disease state[46–48].
Fecal Biomarkers		
Calprotectin	Sensitive marker for intestinal inflammation[32,36,39].	Cannot differentiate CD from UC. Used to monitor disease state[39].
Lactotransferrin	Can distinguish active IBD from inactive IBD and irritable bowel syndrome[49].	Cannot differentiate for CD and UC. However, usable for monitoring disease state[49].

Table 1. Commonly used IBD diagnostic markers, in serum: anti-saccharomyces cerevisiae antibody (ASCA) and perinuclear anti-neutrophil cytoplasmic antibodies (pANCA); and in feces: calprotectin and lactotransferrin[1].

1.2. Rheumatoid Arthritis

RA is a chronic inflammatory joint disease. The disease is characterized by persistent inflammation of the synovial membrane, joint swelling, and chronic inflammation of the synovial joints (Figure 2). The synovial inflammation leads to damage to the articular cartilage and bone erosion, and in many cases ultimately to joint destruction[50,51]. Synovial fluid fills the joint cavity and is responsible for reducing friction within the joint and maintains joint homeostasis. During injury and joint diseases, such as RA, the composition, protein concentration, volume and function of synovial fluid is altered[52].

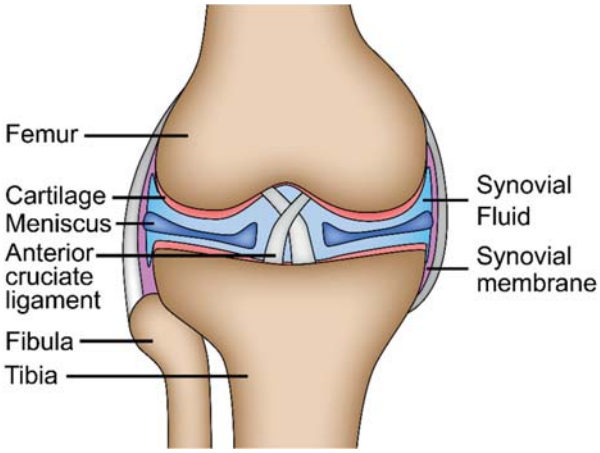


Figure 2. Anterior view of a human knee[3].

Similarly to IBD, RA is more predominant in industrialized countries with a prevalence of 0.5-1.0%, with a higher prevalence in females and increases with age[53,54]. The number of RA patients is increasing, and RA has been identified as one of the most disabling and costly chronic diseases due to lost labor and expenses to the health-care system[54,55].

1.2.1. Citrullination and Rheumatoid Arthritis

As mentioned, smoking is the environmental factor with the highest impact on the risk of developing RA and smoking results in increased abundance of PADs and citrullinated proteins in the lung tissue[14,15]. Citrullination is believed to play a key role in the etiology of RA and citrullinated proteins and peptides have previously been identified in increased amounts in the synovial fluid from

RA patients[56–58]. Autoantibodies targeting specific citrullinated proteins (ACPA), namely vimentin and filaggrin have been described in RA. The latter serve as a disease specific diagnostic marker, and screening for anti-citrullinated peptide antibodies (anti-CCP) is today used as a standard screening method for RA with a sensitivity of 72% and a specificity of 95-99%[50,59–61]. Citrullination is a normal physiological process, mainly in apoptotic cells, and as a result the immune system should not encounter citrullinated proteins. However, it has been hypothesized that an impaired clearing of cell debris during massive cell death could release of PADs and citrullinated proteins to the extracellular environment. This could in turn trigger an immune response, which results in more cell death resulting in a chronic inflammatory state[57]. Whether the hypothesized etiology is accurate remains to be verified. However, ACPAs can be detected in the serum years prior to manifestation of any clinical signs of inflammation in the joints, indicating that the initial triggering of ACPA production might not be located to the joints[62]. The lungs have been suggested as an alternative citrullination site candidate, as smoking as mentioned results in increased citrullination of proteins in the lungs and is the best known environmental risk factor in RA[63]. However, the site(s) of initial immune response triggering remains to be verified.

1.3. Genes, Transcripts, and Proteins

The human genome contains the genes for the expressed human gene products, and the human genetic code contains approximately 20,687 protein coding genes[64]. Proteins are responsible for the majority of the biological functions of the human cells and tissues[65]. Hence, alterations in the genetic code resulting in an altered amino acid sequence of a protein can have severe consequences for the organism. Proteins, therefore, represent an obvious target for studies seeking to discover disease specific markers and explain disease etiologies. During protein synthesis, the DNA code is transcribed into RNA transcripts, and each gene can give rise to several different transcripts. As a result, the approximately 20,687 protein coding genes can result in approximately 100,000 different RNA transcripts[66–68] (Figure 3). Following translation of the RNA into a specific sequence of amino acids, most proteins are covalently modified at several positions[69]. More than 200 different PTMs are known to exist, amongst these is citrullination, and the PTMs are often crucial to the correct function of the protein. The final mature protein products including PTMs and splice variants are termed proteoforms, and it is estimated that the human body can produce more than one million different proteoforms, which constitutes the human proteome[67,70].

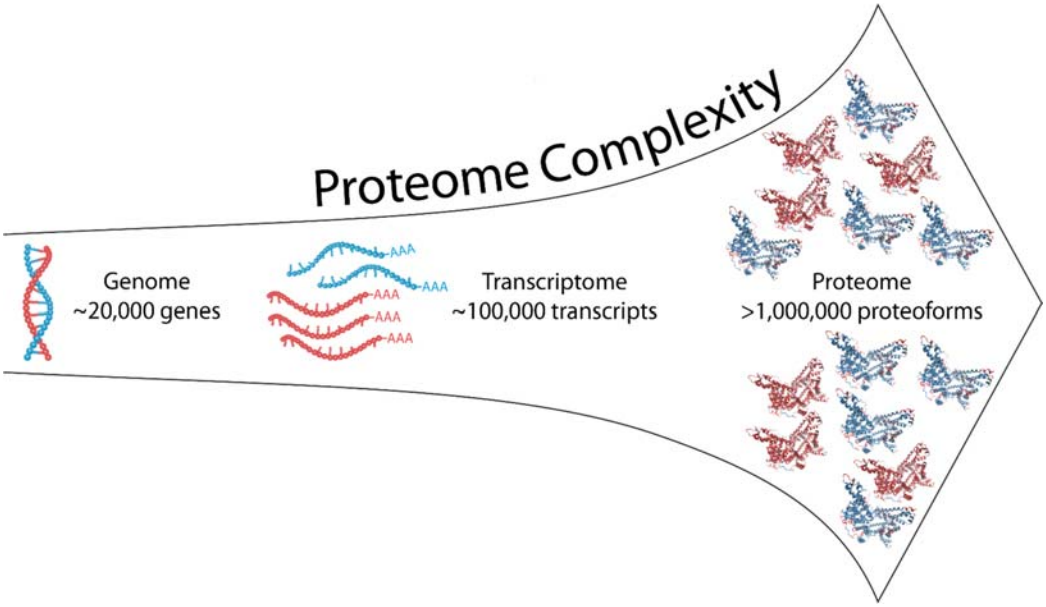


Figure 3. The number of possible gene products increases dramatically from the human genome to the transcriptome, and to the mature protein products (i.e. proteoforms) which in many cases carry PTMs[1].

Information of genetic and gene-products can, in principle, be obtained by studying the genes, the transcripts, or the proteins. Sequencing techniques for studying the DNA code or RNA transcripts have the advantage that entire genomes and transcriptomes can be sequenced with great sensitivity and coverage, in part made possible by the chemical nature of nucleotides and the recent development in sequencing technology[71]. As a result, sequencing studies have increased our knowledge of diseases and biological systems tremendously. Genomic sequencing techniques have greatly increased our knowledge of the IBDs, and several CD and UC loci have been known for more than a decade [66,72,73]. Several cellular IBD specific pathways have been identified, including pathways involved in barrier function, epithelial restitution, microbial defense, immune regulation, reactive oxygen species generation, autophagy, and finally various stress and metabolic pathways associated with cellular homeostasis[19]. Findings on the genomic level, e.g. quantative information, are not always directly reflected in the final protein products, and it is apparent that PTMs only can be studied on the proteome level[71,74]. However, the protein molecules are much more diverse than nucleotides and display a greater chemical variation and a wide dynamic range, which presents a range of problems

for the high-throughput analysis of proteins[71]. Today, no single technique exists capable of identifying all proteoforms of complex systems such as human tissue. However, mass spectrometry-based protein analysis has resulted in the identification of human proteins encoded by 17,294 genes, which constitutes approximately 84% of the total annotated protein-coding genes in humans[70].

1.4. Bottom-up Mass Spectrometry-based Proteomics

In 1994, Wilkins M. defined a proteome as the total protein complement of a genome[75]. Proteomics is the large-scale analysis of proteins, and can cover the study of nearly all expressed proteins of an organism at a given state[76,77]. Bottom-up proteomics strategies involve digesting the proteins in question into peptides. The resulting peptides are analyzed in a mass spectrometer and subsequently reassembled into the originating proteins. In top-down proteomics strategies, the intact proteins are analyzed in the mass spectrometer. However, the high-throughput analysis of highly complex protein mixtures, such as human colonic tissue, using top-down proteomics remain infeasible despite the recent development in the field[78]. This is in contrast to bottom-up strategies where the instrumental and computational developments within the field of MS has improved the speed and sensitivity of the mass spectrometers, which has allowed for the identification of nearly all expressed proteins of complex organisms, such as yeast, within a few hours of measuring time, identifying and (relatively) quantifying several thousand proteins as well as study several PTMs[68,76,79]. Proteomics has thereby become ideal in translational medicine, for biomarker discovery and validation studies, and for the development of new diagnostic tools[80].

1.4.1. Sample Preparation Strategies

The initial digestion of proteins prior to MS analysis is done using specific proteases, which enzymatically cleave all protein products into peptides (Figure 4). Usually the specific protease trypsin is used, which cleaves the C-terminal end of arginine or lysine unless the following C-terminal amino acid is a proline. However, also the protease Lys-C which cleaves the C-terminal end of lysine (or a combination of Lys-C and Trypsin) are widely used. Commonly, a concentration of 1:30-1:50 protease:substrate molar ratio is used, and the enzymatic cleavage is allowed to proceed for approximately 16 hours at 37°C to obtain full digestion of the protein sample[81,82]. To avoid degradation of protein in biological samples by naturally occurring proteases, samples should always be frozen as fast as possible following extraction, and kept on ice when possible until the protease

activity can be inhibited. Otherwise, the subsequent protein identification and quantification might be hampered[83]. The protease activity can be inhibited during sample homogenization by adding strong detergents and/or by adding specific protease inhibiting compounds.

To ensure an efficient proteolysis, homogenization of the sample and denaturation of the proteins is a critical step[84]. For homogenization of tissue samples, commonly grinders, homogenizers with beads, and ultrasonication are used. Protein denaturation is ensured by the use of strong detergents, heating the sample, and by reducing the protein disulfide bonds commonly with dithiothreitol or tris(2-carboxyethyl)phosphine (TCEP). Subsequently, the cysteine reduction is followed by alkylation with iodoacetamide or iodoacetic acid[82]. However, as we and others have demonstrated, the alkylation with iodoacetamide results in several alkylation artifacts, such as carbamidomethylated N-terminals and lysines[3,85]. Therefore, chloroacetamide has been suggested as an alternative alkylating agent with higher specificity toward cysteines[85].

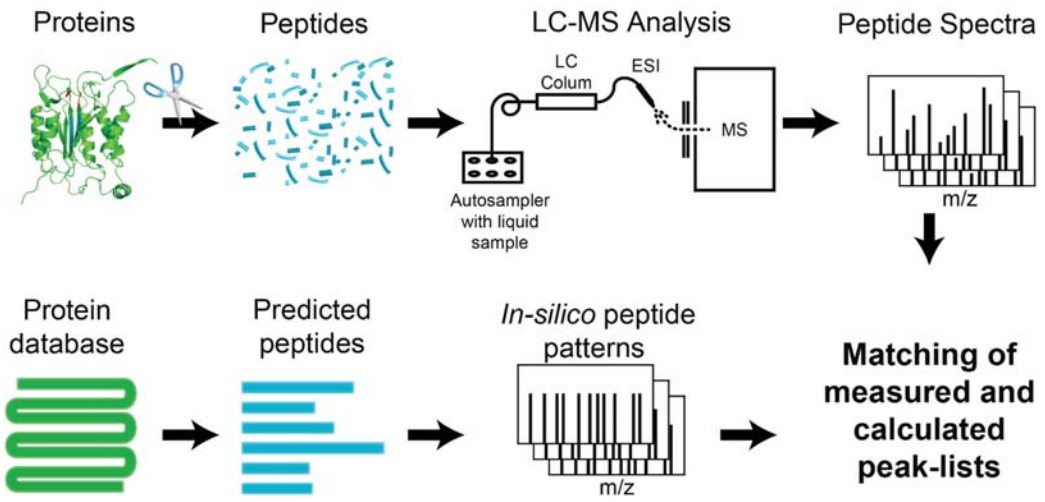


Figure 4. Typical bottom-up MS-based proteomics workflow. The proteins are digested to peptides with specific proteases, and the resulting peptides are separated and analyzed by liquid chromatography (LC)-MS. The measured peptide spectra are matched to an *in silico* generated database, and the peptides and hence proteins in the sample are identified.

Several strategies exist to digest proteins into peptides. The proven and often used in-gel digestion method has been a central protocol for robust protein identification and relative

quantification for decades[86,87]. Prior to enzymatic cleavage of the proteins, the proteins are denatured and kept soluble, often with the powerful detergent sodium dodecyl sulfate (SDS). Many SDS molecules become highly charged after binding to proteins, which prevents the proteins from adsorbing to the container walls. This is in contrast to weak detergents, which might compete with the container walls for protein binding[88]. Following sample homogenization and protein denaturation, the proteins are separated on a polyacrylamide gel in one or two dimensions according to their apparent molecular weight and/or isoelectric point. The separation also facilitates removal of SDS, which inhibits the tryptic digestion and remains MS incompatible, as well as the removal of other interfering substances. By visualizing the protein-spots in the gel with a dye, the protein spots can be quantitated. Spots of interest can be cut from the gel, enzymatically cleaved, and the peptides can subsequently be extracted from the gel and analyzed by MS for protein identification. The method is extremely robust and efficient at removing interfering substances. However, gel-based approaches are suboptimal for samples with high protein complexity and dynamic range[89]. Other considerations is the incomplete recovery of peptides from the gel following digestion, which is concentration dependent and interferes with the subsequent quantitation. Finally, several steps in the in-gel digestion protocol are not readily compatible with throughput techniques, thus reducing repeatability and unbiased protein analysis[3,90,91], which has resulted in a shift toward in-solution digestion strategies.

In solution digestion is an alternative digestion strategy, where all digestion-steps are performed in a liquid phase, in contrast to in a gel matrix. Generally, all digestion steps are performed in a single low-protein binding tube, making the protocol compatible with high-throughput strategies. However, buffer-exchanges are infeasible during the digestion protocol, and can only be implemented after the protein digestion has been completed. SDS is commonly used for protein denaturation at 1-2% concentration, at which the activity of trypsin is inhibited[88]. Furthermore, SDS interferes with the MS analysis if not efficiently removed[88]. The lack of buffer-exchange makes the in solution digestion protocol incompatible with SDS. Urea is usually used as an alternate detergent at 8-10 M urea concentrations, after which the solution can be diluted to 1-4 M urea where the tryptic activity is retained. The protocol can be combined with Lys-C digestion which retains proteolytic activity in high urea concentrations[92]. The lack of buffer exchanges results in an increased final volume during the digestion step and a lower protease concentration in the solution. Furthermore, the presence of chemicals and other interfering small molecules which might interfere with the digestion, decreases the proteolytic activity[84].

Finally, in 2005 a method was published combining many of the advantages of in-gel digestion and in-solution digestion[93]. By performing the digestion on a molecular weight cutoff spin filter, buffer exchanges could be re-implemented to the digestion protocol. During centrifugation, proteins are retained by the filter, while small molecules and solvents can pass through the filter. As a result, the final digestion volume can be kept small, vastly increasing the protease concentration compared to the in-solution protocols. We and others have demonstrated that the protocol can be more repeatable than in-gel digestion and in solution digestion[3,84]. The filter aided sample preparation (FASP) protocol was further developed to include SDS to soluble the proteins, followed by washes with urea, which effectively removed interfering substances prior to digestion[94]. Sodium deoxycholate (SDC) has been identified as an alternate detergent to SDS and urea, and in contrast to most other detergents, SDC was found to increase the activity of trypsin almost five-fold at a 1% concentration[95].

Commonly following proteolytic digestion, detergents, salts, and other contaminants are removed by desalting using solid phase extraction or reversed phased chromatography. The peptides bind to the column and can be washed with acidified water, and the peptides can subsequently be eluted using acidified organic solvents, commonly acetonitrile (ACN). However, SDC also binds effectively to the columns and elute with ACN, so a phase inversion step with an organic solvent such as ethyl acetate is often included in protocols with SDC digestion to facilitate detergent removal[84,96].

1.4.2. Liquid Chromatography Mass Spectrometric Analysis

A tryptic digest of a complex sample, such as a human tissue sample, will result in tens of millions of different peptides. Even though modern MS-systems can sequence up to hundred peptides per second, the overwhelming number of peptides presents a range of technical problems. Separation of the peptides by liquid chromatography (LC) prior to the MS-analysis is often applied (Figure 4). The peptides are loaded onto a column and sequentially eluted over time, using an organic solvent. The LC-system, furthermore, increases the peptide concentration (as the peptides are sequentially eluted), which is a key requirement to facilitate identification of the peptides as MS is concentration dependent. However, the design of the LC-gradients must be optimized to ensure an optimal peptide elution profile, and the finite sequencing speed of the mass spectrometer must be taken into consideration. Ultra-high performance liquid chromatography (UPLC) nanoflow-systems utilize extended columns (>50 cm) to facilitate the MS analysis of complex samples such as human tissue

samples, by increasing the applicable gradient length without compromising the chromatographic peak-width excessively.

Following elution from the analytical column, gas-phase ionized peptides are formed by electrospray ionization (ESI). The peptides are ionized using a high potential (typically 2-4 kV), and a heated drying gas is applied to aid solvent evaporation. In the ESI process, tryptic peptides predominantly become doubly protonated, but higher charge states are also obtained, especially with increasing peptide lengths (roughly one charge per 1.4 kDa). The charge states can be calculated from the isotopic pattern where the isotopic peaks are separated by approximately 1 Da, resulting mainly from the 1.1% naturally occurring carbon-13[78]. Following the ESI process, the peptides are introduced into the mass spectrometer by a vacuum and in the MS manipulated using electrostatic fields. Other ionization techniques exist, such as matrix-assisted laser desorption/ionization (MALDI). However, a review of the different ionization techniques is beyond the scope of this thesis.

When complex samples are analyzed several peptides will elute simultaneously from the LC-column. Usually, a data-dependent mode of acquisition (DDA) is used for the MS analysis, but other strategies exists such as the data-independent acquisition: Sequential window acquisition of all theoretical mass spectra (SWATH). DDA involves measuring the mass-to-charge ratio (m/z) of the eluting precursor ions, from which the molecular weight of the intact peptides can be calculated[69]. Each precursor is, furthermore, isolated and collided with an inert gas, often nitrogen, helium, or argon, which causes the peptide to fragment and the m/z's of the resulting fragments are likewise determined. Predominantly the bonds with the lowest energy are broken using collision-induced dissociation (CID) or higher-energy collisional dissociation (HCD), which mainly are the amide bonds and the methods predominantly produce b- and y- ion series where the charge is retained by the N- and C-terminal, respectively (Figure 5)[78]. Due to the finite speed of the mass spectrometers, commonly only the most intense precursors are analyzed (top n methods) in any given precursor scan. Various strategies exist to perform the MS analysis in a quantative manner, to yield relative protein abundances. The strategies can be divided into labelled approaches where the proteins/peptides are stable isotope labelled prior to the MS analysis, and label-free approaches where a high LC repeatability is obtained and utilized to align different LC-MS analysis. In both approaches, the signal intensity from the MS analysis is assumed to correlate with the peptide (and hence protein) abundance.

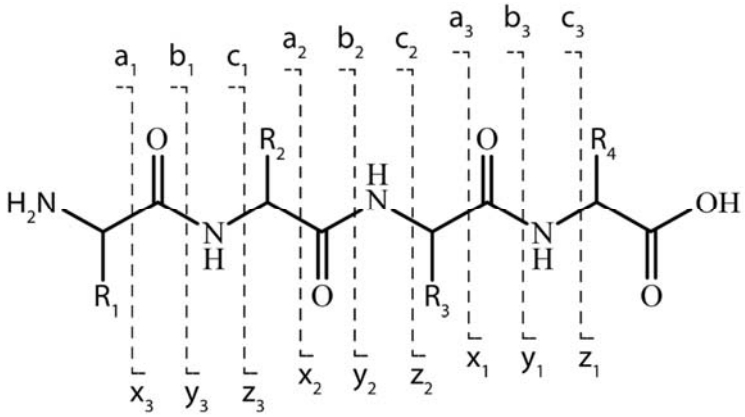


Figure 5. Peptide backbone fragmentation patterns. CID and HCD fragmentation predominantly produces y and b ion series, whereas electron-transfer dissociation (ETD) fragmentation predominantly produce c and z-dot ion series.

1.4.3. Data Processing, Databases and False-Discovery-Rate Estimations

The proteins in the sample are identified post processing by matching the peptide intact masses and measured fragment m/z's against a database, which has been *in silico* generated from a reference database of protein sequences. The reference databases can originate from curaled protein sequences to automatic translated sequences. Transcriptomics is a complementary technique to proteomics, and yields data on which RNA transcripts are present at that specific time[97]. As we have demonstrated, transcriptomics databases can be used to generate tissue-specific databases, and combined with proteomics the likely origin of the identified proteins can be elucidated[3]. The quality of the peptide spectral match (PSM) between fragment m/z' and the *in silico* generated data is assessed and each identification is given a quality score. As many parameters in a LC-MS analysis vary between experiments (the sample complexity, LC-MS systems, databases, etc.) a fixed cutoff for the identification scores cannot be used. Previously, a minimum of two identified peptides have been required for a protein identification[98]. However, a false discovery rate (FDR)-based approach has been found to provide a lower FDR, and the method is today commonly applied[98]. In FDR-based approaches, the entire dataset is searched against a reversed or scrambled "decoy database", which is identical to the real "target database" in terms of size, number of proteins, amino acid distribution etc. The identification score of the true hits in the target database will predominantly be greater than the identification score of the false hits in the decoy database (Figure 6). Based on this, the cutoff for the

identification score is defined as a function of the percentage of decoy hits in the results. Usually 1% FDR is used, but as we and others have demonstrated, the cutoff score should be optimized for the given experiment as the FDR approach is suboptimal for very large or small databases and datasets[3,98]. It is apparent that only proteins and PTMs included in the search can be identified in the analysis. A large database or dataset provides a higher probability that a given mass spectrum will be matched in the decoy database, thereby forcing the applied cutoff to a higher identification score and reducing the number of target hits. On the other hand, as we and others have demonstrated the database and dataset must also not become too small[3,99]. Included PTMs in the search greatly increases the size of the search space, as all possible peptide mass variations are evaluated. It is therefore critical to restrict the search parameters as much as possible.

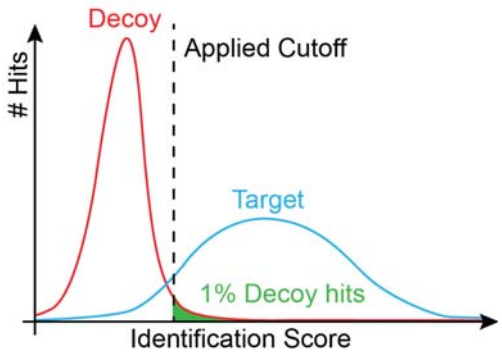


Figure 6. Typical distribution in a target-decoy search. The applied cutoff is defined from the percentage of hits found in the decoy database.

1.5. The Posttranslational Modification Citrullination

Citrullination or deimination is the conversion of the amino acid arginine into the non-standard amino acid citrulline (Figure 7). The reaction is *in vivo* catalyzed by the PAD enzyme family (EC 3.5.3.15)[2]. Each citrullination-event of a protein leads to the loss of one positive charge at physiological pH and results in a monoisotopic mass increase of +0.984016 Da[100]. The loss of positive charge influences the charge distribution, isoelectric point, and hydrogen bond forming abilities of the protein, and the PTM is believed to alter the protein fold, amongst others leading to denaturation in order to promote enzymatic protein degradation[56,101,102]. An example of a citrullinated protein is myelin basic protein, a part of the central nervous system, where it functions as a non-conducting

isolator between the nerve fibers[103,104]. Myelin exists in different charge isomers that differ by the degree of citrullination[102,105].

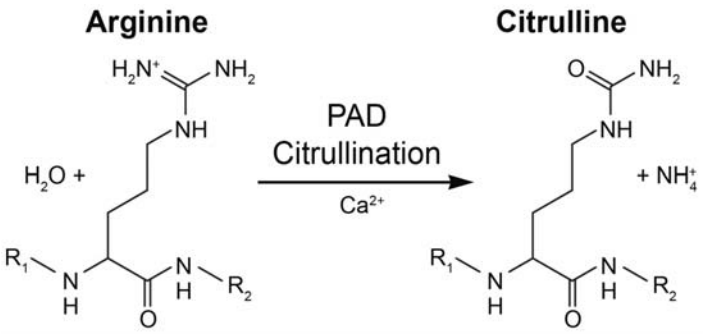


Figure 7. Citrullination of the amino acid arginine into citrulline[2].

1.5.1. Identification of Citrullination with Mass Spectrometry

Several features of citrullination makes it difficult to detect and assign correctly with MS[2,100]. The citrullination mass increase is identical to the mass increase caused by deamidation of asparginine and glutamine, common artefacts likely introduced during exposure to non-physiological pH during the sample preparation[3,106]. Determination of the correct PTM causing the mass increase can therefore be problematic. Furthermore, the mass shift is close to that of a single neutron +1.008665 Da. As a result, if two peptides which are identical apart from a citrullination event, are chromatically coeluting, the m/z-signal from the citrullinated peptide will fall within the isotopic cluster of the non-citrullinated peptide[107]. Due to the typical isolation window used on mass spectrometers (m/z 1-3), both peptides would simultaneously be selected for fragmentation, which most likely would hinder a successful assignment and identification of the PTM[100]. In bottom-up proteomics strategies the protease trypsin is usually the protease of choice for the enzymatical digestion of the proteins prior to analysis[81,82]. As mentioned, trypsin cleaves the C-terminal end of arginine or lysine unless the following C-terminal amino acid is a proline. However, using several synthetic peptides and synovial fluid from a RA patient, we have demonstrated that citrullination unambiguously inhibits tryptic proteolytic activity at the modified arginine (Figure 8), thus resulting in a tryptic missed cleavage[2]. The resulting missed cleavage is in contrast to a deamidation event of asparginine or glutamine, which will not result in a missed cleavage. As a result, a tryptic missed cleavage combined with the correct mass increase and fragment m/z information can be used as a marker for a citrullinated peptide, in

combination with the correct mass increase and fragment masses information. It is worth noting that in the specific case where the C-terminal of a protein is a citrullinated arginine, a missed cleavage will not be reported and using the recommended citrullination identification criteria outlined above such peptides will be dismissed as false positives. Even though such cases can be assumed rare, we have identified such a peptide in the synovial fluid[2]. Thus, if the focus of a given study is a specific protein C-terminal citrullination, these should be manually assessed. Nonetheless, a missed cleavage remains a very significant marker for citrullination when combined with the accurate mass increase, peptide fragment data, and when available a neutral loss[2,108].

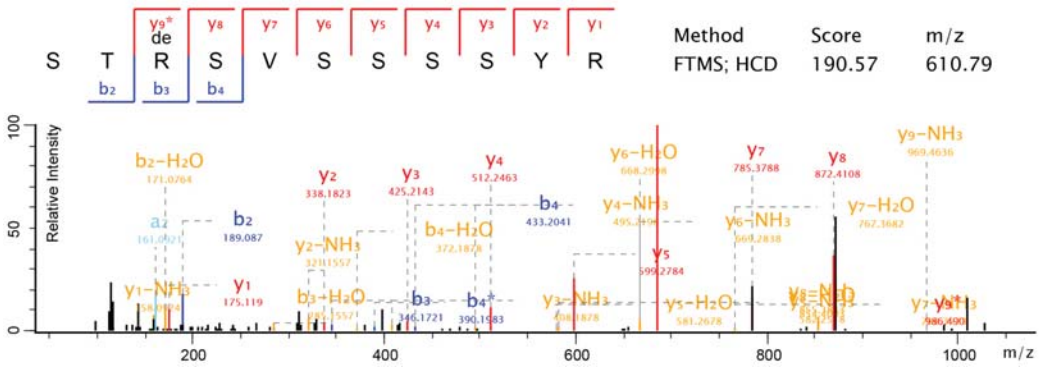


Figure 8. Annotated fragment spectrum of the tryptic digested peptide $STR_{(cit)}SVSSSSYR$[2]. Citrullination inhibits tryptic proteolytic activity at arginine.

The verified inhibition of tryptic proteolytic activity at arginine by citrullination, has the consequence that in a tryptic digest peptides with identical sequences apart from a citrullination will not be present, as the citrullinated peptide lacks a tryptic cleavage site (Figure 9a). Thus, coelution is not a problem for tryptic peptides. However, chromatographic coelution might be an issue for Lys-C digested samples as a Lys-C digestion could result in peptides with identical sequences apart from the citrullination. Citrullination results in an increased hydrophobicity of the modified peptide, and as a result, citrullinated peptides will have a longer retention time on a reversed phase LC column. We have demonstrated that for most investigated peptide pairs, coelution will not occur on standard LC gradients (Figure 9b)[2]. The altered isotopic pattern during peptide coelution has successfully been used to assess the ratio of citrullinated peptide[100]. However, as we have demonstrated, in order to have an overlap of elusion, the LC gradient must be designed very steep, which limits the feasibility of the approach for complex samples such as human tissue samples.

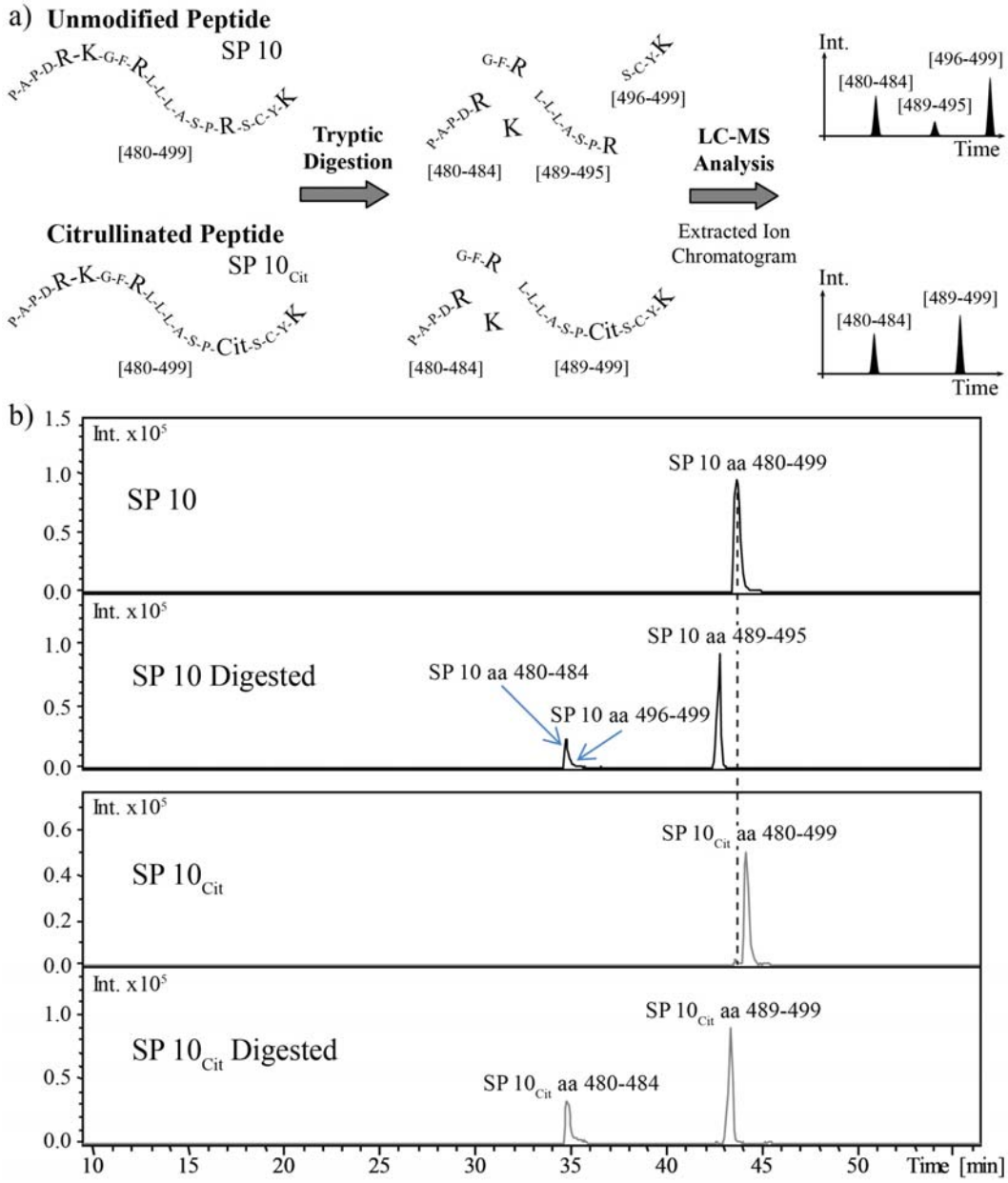


Figure 9. Citrullination inhibits tryptic cleavage at arginine. a) Tryptic digestion of two near-identical peptides SP 10 and SP10$_{cit}$, containing an arginine and a citrulline at position 495 respectively. b) Chromatographic retention time shift between SP 10 and SP10$_{cit}$ on a reversed phase C18 column[2].

1.5.2. Neutral Loss Triggered Precursor Selection

Citrulline residues display a unique neutral loss of HNCO when subjected to CID and HCD, corresponding to 43.0058 Da[108]. The neutral loss is utilized in neutral loss triggered precursor selection MS instrument methods, where the identified MS precursors are enriched with citrullinated peptides[109]. The principle of the method is to conduct a precursor scan followed by an all-ion-fragmentation and a scan of the resulting ion fragments (Figure 10). Any identified mass differences between the two scans corresponding to a predefined neutral loss list, triggers a data-dependent MS scan on the precursor ion. As a result, mostly citrullinated peptides will be chosen for data-dependent fragmentation and identified. The method is dependent on a high signal intensity of the neutral loss (Figure 10, blue circle), as all-ion-fragmentation MS spectra can become highly complex, which can hinder the detection of the neutral loss. Despite optimization of the HCD energy, I was in this PhD project not able to establish a method with sufficient sensitivity for identifying citrullinated peptides in human mucosal biopsies or synovial fluid from rheumatoid arthritis patients (data not shown).

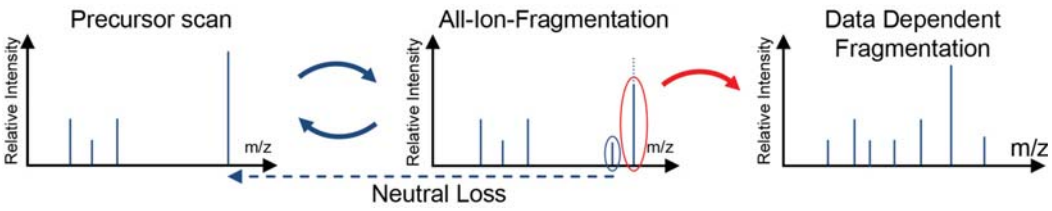


Figure 10. Neutral loss triggered precursor selection. Alterating presursor scans and all-ion-fragmentation scans are performed. When a neutral loss is detected corresponding to a predefined list, a data-dependent fragmentation MS scan is triggered on the precursor.

2. Proteomics Analysis of Colonic Tissue from Ulcerative Colitis and Rheumatoid Arthritis

We are born sterile and within a few days our gut undergo colonization[110,111]. The mature human gut contains more than ten times the number of prokaryotic cells as human cells; these bacteria – the gut microbiota – influence a number of human functions[112]. It is assumed that roughly 100 trillion bacteria from at least 1,000 species are present in the gut, representing an enormous diversity, and millions of years of co-evolution have shaped this human-microbial relationship[113,114]. The metabolic activities of the gut microbiota have major consequences for us, which can be both beneficial and harmful[115]. For instance, a number of microorganisms have been shown to evoke autoimmunity. Additionally, infection with intestinal microbial pathogens such as *Salmonella*, *Shigella* and *Yersinia* can trigger autoimmune reactions in joints and other organs[116]. Furthermore, increased amounts of the sulphate-reducing bacteria *Desulfovibrio* has been associated with UC[112] and segmented filamentous bacteria can induce the IL17-IL23 pathway[117]; a core pathway suggested to be involved in IBD, RA, diabetes mellitus, multiple sclerosis and psoriasis[118–121]. In general, a reduced diversity has been found in the composition of the gut microbiota of IBD patients[122–125], and IBD has been hypothesized to be caused by an impaired interplay between the host and the gut microbiota[112,113,124–131].

The epithelial cells of the intestine form a single-cell epithelial layer with an impressive surface area of 200 m^2, representing the interacting surface of the host-microbial interactions[114]. The intestinal epithelial cells thereby play a key role in defining the mucosal immune response and maintaining intestinal homeostasis. The layer functions as a barrier between the host and the intestinal luminal content, including bacteria and other antigens, while allowing the absorption of water and nutrients[114]. Intercellular junctions, of which the tight junctions are the most critical for maintaining homeostasis, seal the space between adjacent epithelial cells. The tight junctions define the overall barrier function, and play a key role in defining permeability[132]. As a consequence of the critical role, the intestinal epithelia will heal wounds within minutes[133,134]. Besides controlling absorption, the barrier preforms several actions, crucial for maintaining intestinal homeostasis, including secretion of compounds that influence microbial colonization, sensing of the intestinal microenvironment, including beneficial and harmful microorganisms. However, the epithelial cells do not only regulate the intestinal homeostasis, but also responds to signals from the commensal gut microbiota[114,135,136]. The intestinal

immune system is faced with the difficult task of tolerating the complex content of the intestinal lumen, while giving a measured response to threats. The intestinal epithelium is directly involved in the innate immune system, and the barrier is, furthermore, populated by immune cells such as dendritic cells, macrophages, and neutrophils, all essential in maintaining the intestinal homeostasis[19,132,137]. Neutrophils are essential for the innate immunity and are the most abundant leukocytes in plasma. In response to inflammatory stimuli, they migrate from plasma to the effected tissue where they bind and inactivate invading microorganisms. Neutrophils act as the first line of innate immune response to an injury or threat and mediate their effect by releasing enzymes from the granules and produce compounds with antimicrobial potential[138]. In addition to the classical effector molecules, such as reactive oxygen species, the antimicrobial compounds include neutrophil extracellular traps (NETs). NETs are extracellular fibrillar networks composed mainly of DNA, but also contains proteins and other effector molecules some of which have antibody-like properties, from the neutrophil granules. The NETs trap the invading microorganisms and facilitates interaction with other effector molecules that kill the microorganisms[138–141].

As mentioned, proteins are responsible for the majority of the biological functions and cellular behavior in human cells and tissues[65]. In 2001, an NIH group defined a biomarker as "A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention."[142], usable for diagnostics, monitoring disease prognosis and disease monitoring and prediction. Based on the outlined current opinions in the literature, we hypothesized that a proteome study of the intestinal mucosa from UC patients and controls could identify such biomarkers, essential for providing additional insights into the etiology of IBD, as well as function as diagnostic and prognostic disease markers. Motivated by the outlined similarities between IBD and RA, ranging from shared genetic and environmental risk factors as well as an altered gut microbiota composition, we decided to include mucosal biopsies from RA patients, not suffering from IBD. Being a systemic disease, we decided to investigate whether an increased amount of citrullinated protein could be detected in the mucosal tissue, and hypothesized that citrullination might be involved in the etiology of UC as well. PTMs are introduced after translation, thus DNA and mRNA sequencing techniques do not directly identify these modifications, in contrast to proteomics. Fortunately, the technological development in recent years within proteomic research has made proteomics studies of samples as complex as human intestinal tissue feasible following protocol optimization[76,79].

2.1. Experimental Procedures

2.1.1. Collection of Sample Material

The project has been approved by the Danish Data Protection Agency (J.nr. 2011-41-6437) and the regional ethical committee (M-20110151). All patients had given informed consent to participate in the study. Mucosal biopsies from 10 UC patients, 11 RA patients without IBD, and 10 controls without findings were extracted by standard colonscopy with CO_2 at Silkeborg Regional Hospital (Figure 11). Only patients with a well-defined diagnosis were included in the study. Care was taken to only extract biopsies from healthy-looking tissue, and the biopsies were handled on-site by the project group to ensure homogenous sample handling. Biopsies for the proteomics analysis were immediately transferred to individual cryotubes and within 10-20 seconds snap-frozen in liquid nitrogen followed by storage at -140°C. Biopsies for histology were following extraction immediately placed on fixation paper, stabilized in 4% formaldehyde and stored at room temperature until formalin-fixed, paraffin-embedded (FFPE) tissue sample preparation. Serum and buffy coat was collected, and information on medication, diagnosis, and smoking habits was acquired. Furthermore, samples were collected for screening for fecal calprotectin, rheumatoid factor, anti-CCP and antinuclear antibodies (ANA).

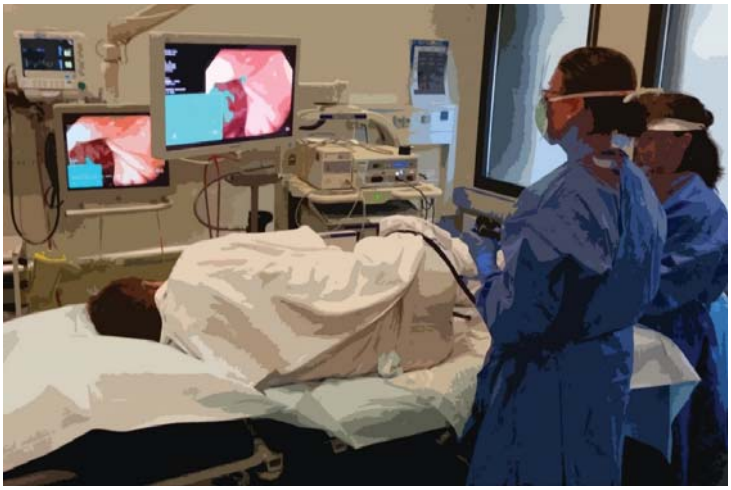


Figure 11. Extraction of biopsies by standard colonscopy. Picture from the present project.

2.1.2. Histology

The colonic biopsies were saturated in paraffin and sliced. The slices were incubated in a water bath at 60°C for 20 min, rinsed three times with water to remove paraffin residues, dipped 10 times in 99% ethanol, dipped 20 times in 96% ethanol and rinsed with water. The samples were incubated at room temperature in hematoxylin Mayer's (Sigma-Aldrich, St. Louis, MO, USA) for 15 min. and rinsed with two changes of water. Samples were dipped twice in 0.25% ammonium hydroxide, and then twice in water, dipped 10-20 times in an eosin solution, then 20-30 times in 96% ethanol and 20-30 times in 99% ethanol, before cleaning by dipping 20-30 times in tissue clear (Sakura, AJ Alphen aan den Rijn, The Netherlands).

2.1.3. Protein Digestion

Intestinal biopsies (approximately 2-3 mm^3) were kept on ice when possible to avoid unwanted protein degradation/modification due to enzyme activity. The frozen biopsy was transferred to an impact resistant tube with five steel-beads on a scale, and the wet weight was determined. The tube was immediately placed on ice and 0.5 mL cold lysis buffer was added (5% SDC, 50 mM triethylammonium bicarbonate, pH 8.5). The samples were homogenized using a Precellys 24 homogenizer (Berting Technologies, Rockville, MD, USA) at 6,000 rpm for 20 sec a total of 3 times. Between the runs, the samples were cooled on ice. The samples were sedimented at 8,000 g for 30 sec, and inspected for large debris. Total protein concentration was determined for normalization of sample material for digestion on four of the samples using a standard bicinchoninic acid assay (BCA) with bovine serum albumin (BSA) as standard. A total of 20 µL sample was analyzed per well, and a dilution series (1, 1/2, and 1/4) with lysis buffer was loaded, all in triplicates, for accurate protein concentration measurements. The detected protein abundances were held against absorbance at 280 nm (A280) of 1 µL analyzed in triplicates using a NanoDrop 1000 UV-Vis Spectrophotometer (Thermo Scientific, Waltham, MA, USA), calibrated with lysis buffer. A coefficient to transfer from A280 to BCA measurements was calculated, and A280 measurements were used henceforth.

100 µg protein was digested using the FASP protocol with 0.5% sodium deoxycholate and ethyl acetate surfactant removal[95,143]. Following trypsin digestion, 50 µg peptide material was desalted with TARGA C18 columns according to the manufacturer's protocol (Nest Group, Southborough, MA, USA). The eluent was dried down in a vacuum centrifuge overnight and stored at -80°C until time of analysis. The dry peptide product was mixed with 30 µL resuspension buffer (2% acetonitrile (ACN), 0.1% formic acid (FA)), vortexed, sonicated for 2 min, and briefly spun down.

2.1.4. Liquid Chromatography Mass Spectrometry Analysis

The loaded samples amounts were normalized using A280 on a NanoDrop 1000 (Thermo Scientific, Wilmington, DE, USA), calibrated with resuspension buffer. Five µg total protein was analyzed per LC-MS run, diluted with resuspension buffer to a total volume of 10 µL. Research randomizer[144] was used to ensure a random order of biopsy analysis, and all biopsies were analyzed in triplicates. The triplicates were analyzed consecutively (i.e. not randomized), and followed by a quality control with BSA to the ensure system performance, retention-time stability, and ensure minimize sample cross contamination between different biopsy analysis. The autosampler tray was kept at 5°C and no samples resided longer than 48 hrs in the auto sampler prior to LC-MS analysis.

The samples were analyzed using a UPLC-nanoESI MS/MS setup with an UltiMate™ 3000 UPLC system upgraded with a RSLC nanopump module. The system was coupled online with an emitter for nanospray ionization (New objective picotip 360-20-10) to a Q Exactive mass spectrometer (Thermo Scientific, Waltham, USA). Ten µL sample (normalized to five µg) was loaded onto a C18 reversed phase column (Dionex Acclaim PepMap RSLC C18, 2 µm, 100 Å, 75 µm x 2 cm) and separated using a C18 reversed phase column (Dionex Acclaim PepMap RSLC C18, 2 µm, 75 Å, 75 µm x 50 cm) at 40°C. The sample was eluted with a gradient of 96% solvent A (0.1% FA) and 4% solvent B (0.1% FA in ACN), which was increased to 8% solvent B on a 5 minutes ramp gradient and subsequently to 30% solvent B on a 225 minutes ramp gradient, at a constant flow rate of 300 nL/min. The mass spectrometer was operated in positive mode in DDA, selecting up to 12 precursor ions based on highest intensity for HCD fragmenting. Fragmented ions were dynamically excluded for 30 sec. MS1 resolution was 70,000 and MS/MS resolution 17,500, both with a maximum injection time of 70 ms and 1E6 and 1E5 automatic gain control (AGC) targets respectively. The scan range was set to m/z 400-1200, with an isolation window of m/z 1.6, and a normalized collision energy of 27. Peptide match was set to preferred, and exclude isotopes was set to on. Undetermined, +1 and >+8 charge states were excluded, and an underfill ratio of 3.0% was used, yielding an intensity threshold of 8.6E4. Apex triggering and lock mass was set to off.

2.1.5. Proteomics Data Analysis

The data-files (.raw) were searched using MaxQuant 1.4.1.2 against all reviewed *Homo sapiens* Uniprot proteins (downloaded 08/10/2013, containing 20277 entries). All standard settings were employed with carbamidomethyl (C) as a static modification and protein N-terminal acetylation, deamidation (NQ) and citrullination (R) were included as variable modifications. Citrullination was

not allowed to be present on the C-terminal in order to reduce the number of false positives when assessing citrullinated peptides[2]. All proteins and peptides are reported below a 1% FDR cutoff, to ensure only high-confidence protein identifications.

Initially, all reverse hits and proteins tagged as contaminants were removed from further analysis, and the data was log2-transformed. Two unique peptides or more were required for a protein identification and quantitation to ensure high-confidence quantitation. Furthermore, the proteins should be quantifiable in at least half of the subjects in at least one group (UC, RA, or control group). The protein abundances were investigated replicate-wise, by scatterplots of all protein quantifications. Any replicates within the analysis of a given biopsy with a Pearson correlation coefficient < 0.95 was removed from further analysis. In this way, only highly reproducible LC-MS runs were included in the analysis. The replicates were combined by the median, to yield biopsy (subject) protein abundances. Solemnly for the purpose of conducting principal component analysis (PCA), missing values were replaced with values from a normal distribution (width 0.3 and down shift 1.8), calculated separately for each subject. MS measurements were grouped according to the disease state, and two-sided t-tests were performed, separately calculated for UC and control group, and RA and control group. The statistical tests were corrected for multiple hypothesis testing using permutation based FDR (FDR=0.15, S0=1 and 250 randomizations)[145,146]. Likely due to a high degree of biological diversity and number of proteins quantified, the FDR had to be adjusted to 15%. Besides this, all standard settings were employed. Semi-absolute protein abundances are represented as protein intensity-based absolute quantitation values (iBAQ)[145].

To analyze tentative interactions between the significantly changing proteins, we used the STRING 9.1 software that builds functional protein-association networks based on available experimental data[147]. Gene ontology annotations were imported using STRAP v1.5[148] and Venn diagrams were created with BioVenn[149] and Venny[150] for the data analysis.

2.1.6. Citrullination Analysis

Four filters were applied for a positively identified citrullinated peptide: 1) less than 1% peptide FDR; 2) A missed cleavage after arginine was required[2]; 3) A localization score of 0.75 or greater was required for each individual citrullination localization[151–154]; 4) the citrullinated peptide should be quantifiable in at least half of the subjects in any group (UC, RA, or control group).

The intensities of the citrullinated peptides were divided with the summed intensity of all detected peptides for that given LC-MS run. In this way, we normalized the detected citrullinated peptide intensities to account for possible differences in loaded peptide material, system performance etc. The technical replicates were combined into subjects by taking the median of the intensity. Peptides displaying significant change in intensity between the groups, were calculated by two-sample t-tests using Perseus with both sides and permutation-based multiple hypothesis testing correction using permutation based FDR (FDR=0.05, S0=1, 250 randomizations)[145,146]. The calculations were performed separately for the UC and control group, and RA and control group respectively. Plots were created with RStudio version 0.98.507 and BoxPlotR[155].

2.2. Results & Discussion

2.2.1. Patient Cohort and Biopsy Evaluation

Biological material was successfully collected from 10 UC patients, 11 RA patients, and 10 controls with no findings. The mucosal biopsies were in all cases obtained from non-inflamed areas at approximately 40 cm from the anus (Figure 12a). A well-defined colonic biopsy extraction-depth is important, as protein abundance changes along the length of the colon have been described. In the study, most proteins were found stable in abundance. However, proteins involved in bacterial sensing, cation-transport and O-glycosylation were decreased in the distal direction, whereas proteins involved in bacterial defense and anion-transport were increased[156]. All extracted biopsies visually appeared similar in size and color, and the average wet-weight of the biopsies was measured to be 4.9 mg ±1.3 mg. The histological analysis of the hematoxylin and eosin (H&E) stained formalin fixed paraffin embedded (FFPE) colonic biopsy revealed no apparent tissue changes like mucosal distortion caused by the UC or RA disease state, and all investigated biopsies seemed similar with preservation of the crypt architecture (Figure 12b). However, a x40 magnification revealed an increased density of cell nuclei in the UC biopsies (Figure 12c), which could likely be neutrophils. This theory is supported by the apparent neutrophil infiltration of the crypt epithelium, which is commonly seen in UC[157,158].

Initially, the protein concentration of the homogenized biopsy lysate was determined with BCA and protein A280 using four samples. A highly reproducible mean coefficient of 0.47 (STD 0.01) BCA to A280 was calculated. Due to the high amount of sample required for accurate BCA protein concentration measurements compared to A280 protein concentration measurements (53 µL and 3 µL respectively), A280 measurements were adjusted and used henceforth. Based on the protein concentration measurements, the homogenization of the samples was found to be highly reproducible. Of the biopsy wet weight, 10% (STD 2%) could be detected as soluble proteins validating the chosen lysis buffer and homogenization method.

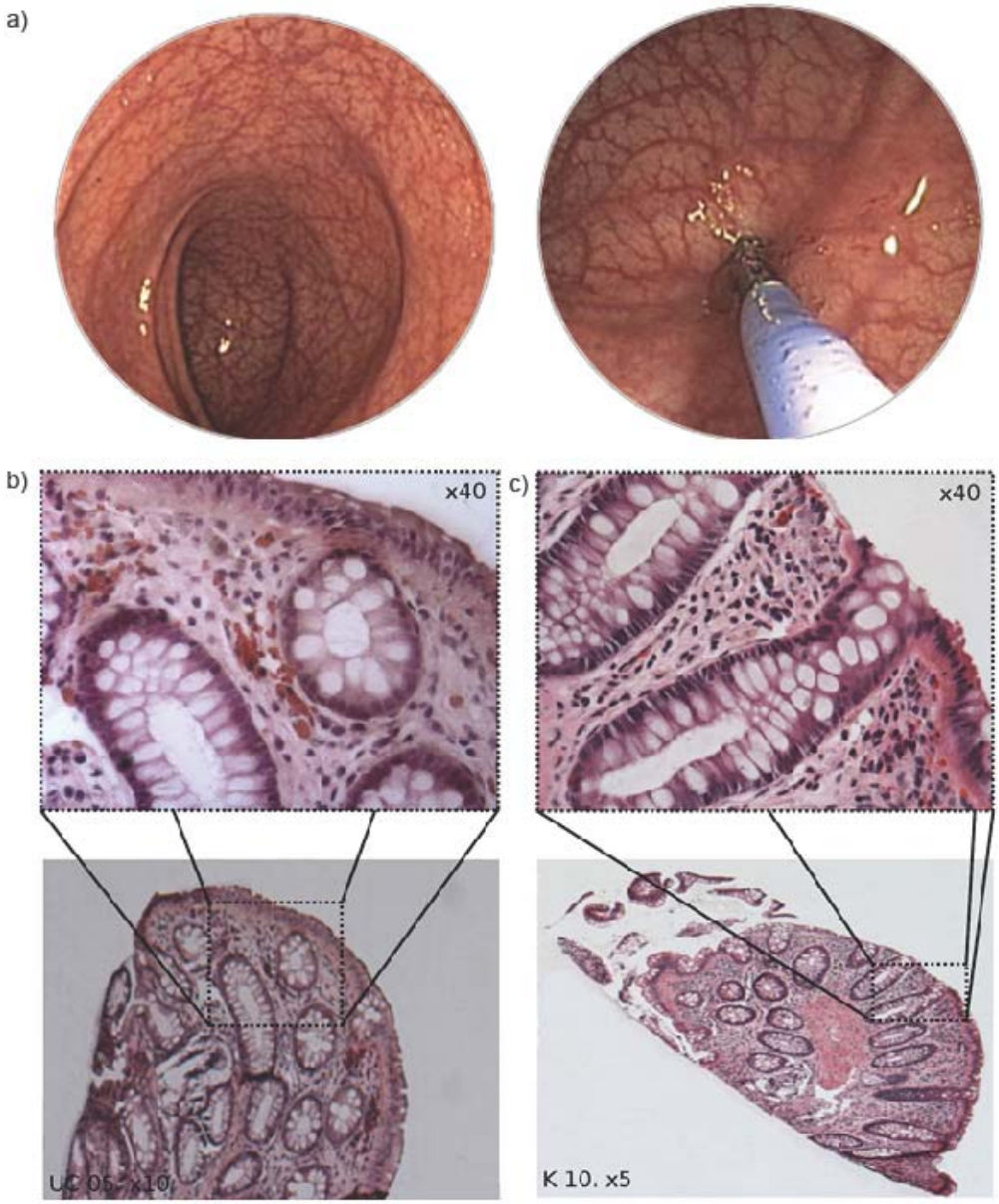


Figure 12. Biopsy extraction and analysis. a) Picture taken during our colonic biopsy extraction, demonstrating a typical view of the biopsy extraction. b) H&E stained FFPE slice analysis colonic biopsy from control group. c) Colonic biopsy from UC group. The crypt architecture is preserved in the UC group. However, an increased number of stained cell nuclei is visible, which could be PMNs.

2.2.2. Proteomic Data Validation

All biopsies were analyzed in three technical replicates, and the technical replicates were analyzed consecutively, in contrast to the randomized biopsies. We calculated the Pearson correlation coefficient pairwise between the triplicates for all subjects, and removed one replicate from three different subjects (3 of 93 total LC-MS runs) with a coefficient less than 0.95 from further analysis. The generally high Pearson correlation coefficients between the replicates points to a high degree of technical repeatability of the method, i.e. a low variance of the calculated protein abundances when the same biopsy is analyzed repeatedly. A PCA of all quantifiable proteins was conducted prior to combination of the replicates (Figure 13). The plot of PCA component 1 and component 2 revealed groupings of the technical replicates, as expected for a low technical variance relative to the inter-biopsy variation. To ensure that the clustering of the technical replicates of the patients were not caused by the sequential LC-MS analysis of the replicates, in contrast to the subjects which were randomized, we investigated whether the clusters of replicates were dependent on the time of analysis. No pattern could be observed (data not shown), which indicates that the time of biopsy LC-MS analysis has not affected the measurements to a detectable extend. The variance in protein abundances between different biopsies, therefore, seems much larger than the technical variance.

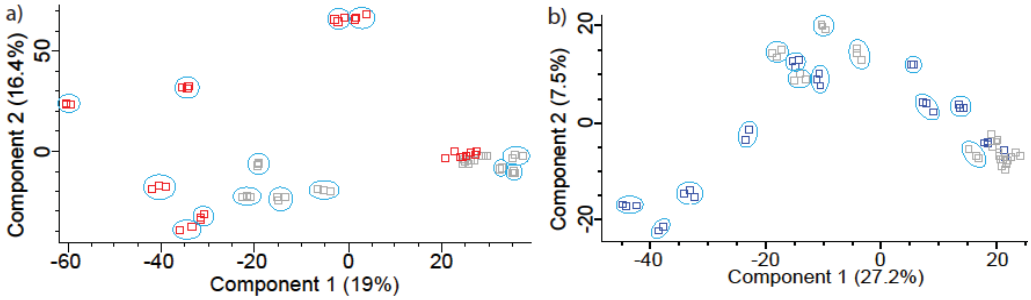


Figure 13. Scores plot of component 1 and component 2 in the PCA of protein abundances. Technical replicates (encircled in blue) group together in all cases; a) PCA scores plot of UC samples (red) and control (gray); b) PCA scores plot of RA samples (blue) and control (gray).

2.2.3. Mucosal Protein Abundance Analysis in Inflammatory Diseases

The technical replicates were combined subject-wise by taking the median. After the filtering steps 5,444 proteins remained (1% FDR) meeting our requirements for relative quantification. In contrast to the high degree of tissue similarity observed in the histological analysis, a large degree of

diversity was observed on the protein level. The observed small technical variance validates the overall method and analysis, and points to differences between the biopsies or the subjects. A permutation-based FDR level of 15% was chosen for identifying proteins with a statistically significant abundance change between patients and controls, to give sufficient input in the subsequent analysis. The need for the relatively high FDR can be explained by generally large differences between the biopsies. No cell isolation was employed following biopsy extraction. As a result, any variation in the biopsy-tong penetration depth will have resulted in an altered cellular composition of the extracted biopsies. Also due to interpersonal differences the protein composition can be expected to vary between the biopsies. At the applied cutoff, 46 proteins demonstrated a statistical significant abundance change between the biopsies from the UC patients and the biopsies from the control group. Of the 46 proteins, 33 proteins were more abundant in the UC group and 13 proteins less abundant (Table 2). None of the 46 proteins displayed a mean fold-change less than 2.8 between the groups. Interestingly, no significantly changing proteins could be detected between the RA group and the control group. Observing only the number of significantly changing proteins, it appears that the intestinal mucosal tissue of RA patients is much more similar to that of the control group than that of UC patients. This finding is in agreement with the histological analysis. However, non-statistically significant protein changes were detected, hence it is possible that a study including a larger cohort could detect statistically significant protein changes.

Fold Change	**Protein Description**	**Uniprot ID**	**Unique peptides**
219.2	*Lactotransferrin	P02788	49
92.1	Matrix metalloproteinase-9	P14780	30
63.3	*Myeloperoxidase	P05164	41
31.4	Myeloblastin	P24158	7
31.3	*Neutrophil elastase	P08246	11
18.0	Dual oxidase 2	Q9NRD8	42
16.7	Integrin alpha-M	P11215	33
16.3	*Protein S100-A9	P06702	10
13.7	Neutrophil gelatinase-associated lipocalin	P80188	12
11.8	*Protein S100-A12	P80511	5
10.6	Neutrophil collagenase	P22894	14
9.9	*Neutrophil defensin 3	P59666	5
9.4	Ig gamma Fc region receptor III-A	P08637	2
9.2	Hexokinase-3	P52790	18

Fold Change	Protein Description	Uniprot ID	Unique peptides
8.7	Peptidoglycan recognition protein 1	O75594	5
8.5	Chitinase-3-like protein 1	P36222	14
7.8	Regenerating islet-derived protein 3-alpha	Q06141	4
7.3	Dual oxidase maturation factor 2	Q1HG44	4
6.6	Eosinophil lysophospholipase	Q05315	11
6.6	Nitric oxide synthase, inducible	P35228	41
6.2	Arachidonate 15-lipoxygenase	P16050	25
5.9	*Eosinophil cationic protein	P12724	10
5.7	Mast cell-expressed membrane protein 1	Q8IX19	3
5.1	Non-secretory ribonuclease	P10153	8
5.0	Eosinophil peroxidase	P11678	44
4.5	*Myeloid cell nuclear differentiation antigen	P41218	17
4.4	Ig kappa chain V-II region RPMI 6410	P06310	3
4.3	Rac exchanger 1 protein	Q8TCU6	8
4.0	Cystatin-A	P01040	2
4.0	Derlin-3	Q96Q80	3
3.9	Transmembrane protein 179B	Q7Z7N9	2
3.8	Protein unc-13 homolog D	Q70J99	16
3.5	*Cathepsin G	P08311	16
-2.8	Myristoylated alanine-rich C-kinase substrate	P29966	11
-3.1	Alcohol dehydrogenase 1C	P00326	10
-3.4	HLA class II histocompatibility antigen, DRB1-11 beta chain	P20039	2
-3.5	Peptide YY	P10082	3
-3.7	EGF-containing fibulin-like extracellular matrix protein 1	Q12805	15
-3.7	Cytochrome P450 4F11	Q9HBI6	3
-3.9	Elastin	P15502	6
-4.0	Hyaluronan and proteoglycan link protein 1	P10915	13
-4.1	NADH dehydrogenase 1 alpha subcomplex subunit 11	Q86Y39	7
-4.4	Phosphoenolpyruvate carboxykinase	P35558	34
-4.5	F-box only protein 38	Q6PIJ6	6
-4.9	Trefoil factor 3	Q07654	3
-5.1	WAP four-disulfide core domain protein 2	Q14508	4

Table 2. List of the 46 proteins displaying a statistical significant change between the UC group and the control group. * proteins have been associated with NETs[141].

Of the 46 significantly changing proteins between the UC group and the controls, S100A9 calgranulin-B/Calprotectin, neutrophil elastase, myeloperoxidase, alpha defensin 1, cathepsin G have previously been found with increased abundance in colonic biopsies from UC patients[159]. Also, protein S100-A9 and protein S100-A12 has been found to be elevated in the serum of children with IBD[160]. These findings validate the chosen method. An overabundance of the enzyme indoleamine-2.3-dioxygene has been found in intestinal mucosal tissue from CD and UC patients compared to normal mucosa, hypothesizing an involvement of the Kynurenine pathway of tryptophan metabolism in the IBDs[161]. The protein was in our study detected displaying an 1.8 fold increase in UC patients, but with no statistical significance, indicating that a larger study cohort could have identified a higher number of statistically significantly changed proteins. Isolated epithelial cells and not whole intestinal mucosal biopsies were analyzed in the mentioned study using 2D-PAGE protein separation and quantitation followed by MS identification. Indoleamine-2.3-dioxygene activity has been found to be essential in dendritic cells to induce co-cultured T cell apoptosis[161] and thus might not be present in vastly increased amounts in well-treated UC patients. However, it might be that the high degree of biological diversity encountered could be reduced by analyzing an isolated/enriched cell group.

An analysis of the measured iBAQ values of previously identified significantly changing proteins between mucosal biopsies from UC patients and controls, revealed that many of the previously known biomarkers are in the upper half of the iBAQ range (Figure 14). This is in contrast to this study, where the identified significantly changing proteins also are present in the lower half iBAQ area. This points to an increased sensitivity of the methods used in this study.

The protein with the largest abundance change between the UC and control group is lactotransferrin, which has been found to be >200 times more abundant in the UC group (Table 2). Lactotransferrin is used as a fecal biomarker for intestinal inflammation, and is involved in the mucosal innate immune system response. It is an iron-binding glycoprotein expressed by activated neutrophils[164] and is released by the injured tissue. The protein has been found to modulate inflammation and act in the defense against infections as a part of the innate immune system[165]. It is resistant to degradation and proteolysis, and unaffected by freeze thaw cycles, making it a useful biomarker for inflammation[39]. Taken together with the top 10 proteins more abundant in the UC group: Matrix metalloproteinase-9, myeloperoxidase, the leukocyte serine protease myeloblastin, neutrophil elastase, dual oxidase 2, integrin alpha-M, protein S100-A9 and A12, and neutrophil gelatinase-associated lipocalin, this points to a highly activated innate immune system. This is interesting as all UC patients were in remission. However, the finding is in agreement with the

histological analysis, which found an increased abundance of cells, which likely was neutrophils. Both results indicates an activated immune system in the UC colonic tissue, and points a subclinical chronic inflammation in the tissue.

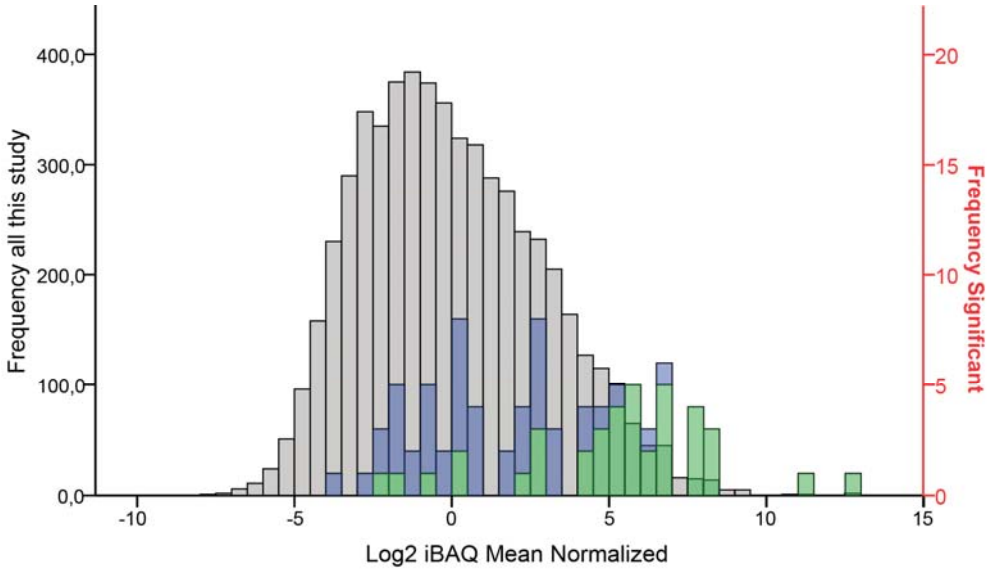


Figure 14. Protein abundances represented by iBAQ values of all quantifiable proteins in this study (gray), all significantly changing proteins in this study (blue), and identified biomarkers in other proteomics studies comparing UC to healthy mucosal tissue, where a corresponding iBAQ value could be found in the present study[159,162,163].

2.2.4. Host-Microbial Interactions and Inflammatory Bowel Disease

Several studies have found alterations in the gut microbial composition during IBD, and it has been hypothesized that altered host-microbial interactions are central in the IBD etiology. A large genetic study found 163 IBD loci, many of which were involved in the host defense against infections[20]. Taken together these findings support that IBD is caused by an inappropriate immune response toward the gut microbiota.

As mentioned, several of the statistically significantly more abundant proteins in the colonic tissue of UC patients compared to controls, are directly involved in the mucosal innate immune system response. A closer analysis revealed that several of the proteins, furthermore, have been

associated with NETs (Table 2). Besides the statistically significantly more abundant proteins, other NET associated proteins were identified with statistically insignificant increased mean abundance fold change (Figure 15). However, not all proteins known to be associated with NETs were found in increased amounts in the mucosal tissue from UC patients, and several of the identified NET-proteins are involved in the normal innate immune system response. Therefore, if the increased presence of these proteins is an indication of an increased abundance of NETs or merely an activated innate immune system remains unclear.

Based on the results, it is apparent that even though the UC patients were well treated and the tissue healthy looking, a chronic inflammatory condition is present in the colon tissue. The results point to an activated innate immune system in the mucosal tissue of the UC patients, and an obvious target for the activated immune system is the gut microbiota. IBD has previously been associated with an altered gut microbiota composition, and it has been hypothesized that IBD is caused by a mucosal immune system response toward the gut microbiota[20]. Our findings, thereby, supports the current knowledge.

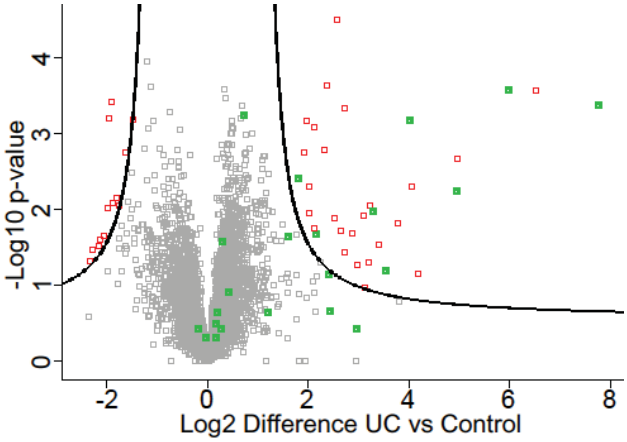


Figure 15. Average protein abundance changes between UC patient and controls (positive X-axis values indicate more abundant protein in UC tissue) plotted against the p-value for the abundance change. The applied significance cutoff (black line) and significantly changing proteins (red) are illustrated. Green indicates that the protein is associated to NETs. A total of 46 proteins were found to be significantly changed of which 33 proteins were more abundant in the UC group. Nine of these have been found to be associated with NETs[140,141].

2.2.5. Citrullinated Peptides

Finally, we conducted an analysis of all identified citrullinated peptides. A total of 231 citrullinated peptides were identified meeting the strict identification criteria. One citrullinated peptide: 446-TVETR$_{(cit)}$DGQVINETSQHHDDLE-466, the C-terminal end of the cytoskeletal protein vimentin, demonstrated a statistically significant intensity-change between the RA and control group at 5% FDR. No peptides with a statistically significant abundance change between the mucosal colonic biopsies from the UC and the control group were found. The peptide was identified with high confidence by MSMS data (Figure 16).

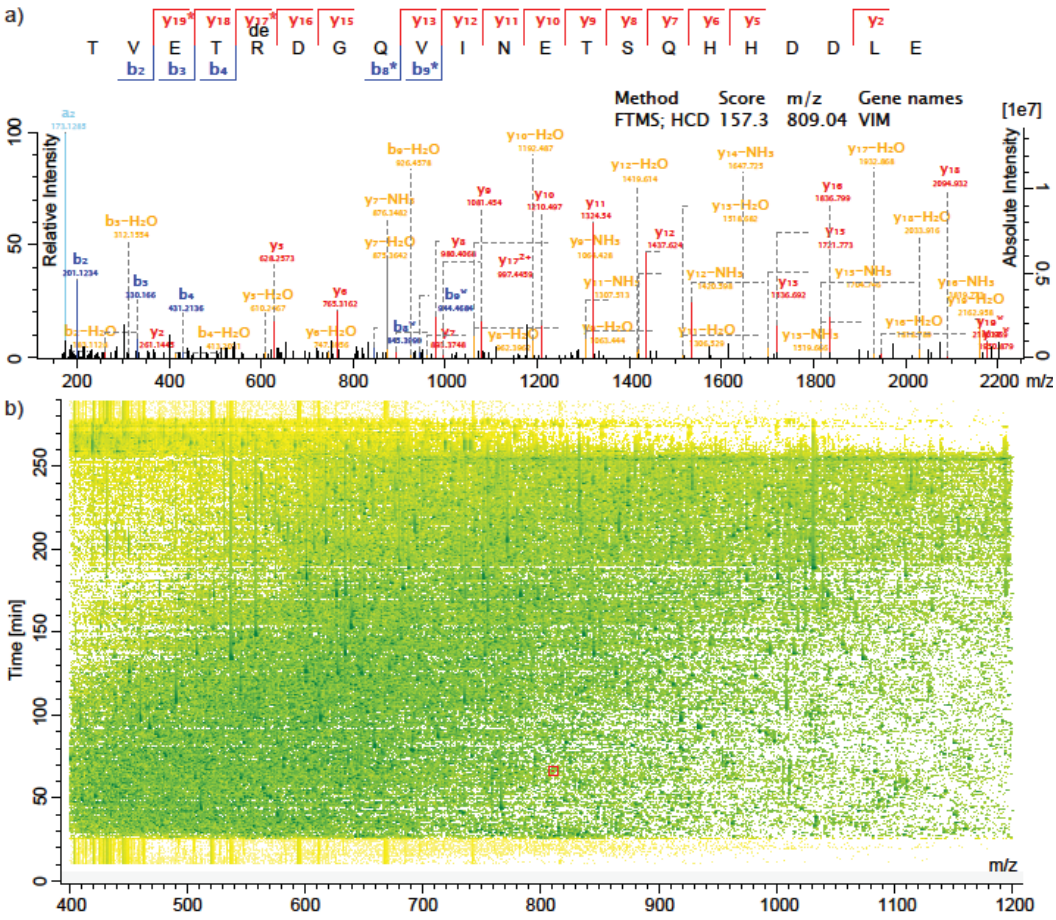


Figure 16. Identification of the citrullinated peptide TVETR(cit)DGQVINETSQHHDDLE. a) One of the annotated MSMS spectra identifying the peptide. b) TIC MS chromatogram of eluting peptides, demonstrating the sample complexity. Red square marks the elusion of the peptide.

The boxplots (Figure 17) revealed that the majority of the detected difference is caused by four RA subjects. One could speculate that this represents a subgroup of the RA patients. However, the recorded medical history and drug usage of these individuals could not explain the grouping. The identification of citrullinated vimentin out of the more than 5,000 identified proteins is striking, as antibodies of citrullinated vimentin previously have been associated with RA[2,166]. Vimentin has been identified as one of the first cellular proteins to become citrullinated in the RA etiology. Indeed, it has been hypothesized that an impaired clearing of active PAD enzymes released during cell death could result in the citrullination of cellular proteins, which in turn serve as the first citrullinated antigens to the immune system[57]. C-terminally citrullinated vimentin has previously been associated with RA, and a recent study has found the same citrullinated peptide increased in the lung tissue of RA patients, hypothesizing that the lungs might be the initial place of ACPA production triggering[63]. The finding of this peptide in the intestinal tissue points to the systemic nature of the RA disease, as well as underlines the involvement of citrullinated vimentin in the RA etiology.

It is possible that a study focusing exclusively on citrullination will identify more citrullinated peptides than this present study. Highly specific methods for the enrichment of citrullinated peptides exists[167–170]. However, the lack of sensitive methods remain a restriction for such studies[2].

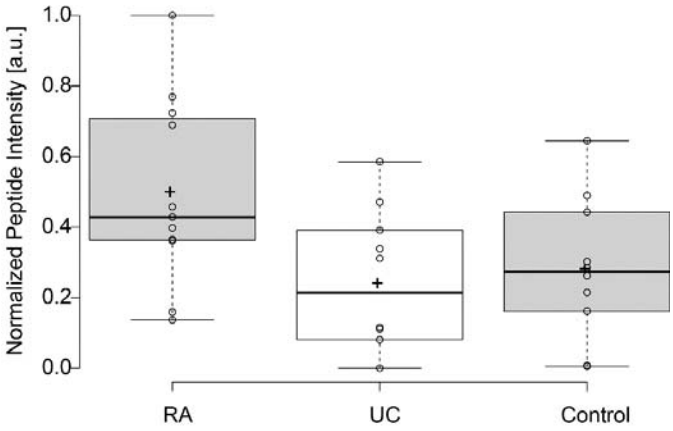


Figure 17. Turkey style boxplots of the TIC normalized intensity of the peptide citrullinated peptide TVETR(cit)DGQVINETSQHHDDLE which was found in overabundance in mucosal biopsies in RA tissue. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; crosses represent sample means; data points are plotted as open circles[155].

2.3. Conclusion

We successfully collected mucosal biopsies from ten UC patients in remission, 11 well-treated RA patients and 10 controls. Based on a histological analysis of the H&E stained FFPE colonic biopsy, the mucosal tissue from RA patients appear indifferent from that of healthy individuals. However, an increased density of cell nuclei was found in the mucosal of the UC patients in remission, which likely is a neutrophil infiltration of the crypt epithelium.

The biopsies were analyzed by LC-MS, and we found a high degree of technical repeatability of the optimized sample preparation method in terms of protein yields. A total of 5,444 proteins met our strict quantitation criteria, and through a quantitative analysis we identified 46 proteins with a statistically significant abundance change between the UC and control group. No proteins were found, demonstrating a statistically significant abundance change between the RA and the control group. By comparing the calculated protein intensities of the identified statistically changing proteins and previously identified UC biomarkers, the superior sensitivity of the current study was demonstrated. Most known UC biomarkers were found in the high abundance region, whereas several proteins found to be statistically significantly changing in this study were found with low abundance. Several new proteins previously not known to be in increased in UC were identified, and several known UC biomarker candidates were validated. Many of the 46 proteins, e.g. lactotransferrin, myeloperoxidase, and dual oxidase 2, are directly involved in the mucosal innate immune system response to microorganisms. Furthermore, several of the more abundant proteins have been associated with NETs points to an increased presence, both indicating an activated innate immune system in the UC colon likely targeting the gut microbiota. The findings in this study thereby suggests that UC is caused by a complex interplay between the host immune system and the gut microbiota.

Finally, using our optimized method for identifying citrullinated peptides, we identified one citrullinated peptide from the C-terminal part of vimentin, demonstrating a statistically significant abundance increase between the RA and the control group. The peptide has previously been identified in lung-tissue from newly diagnosed RA patients hypothesizing the lungs to be the initial site of RA antigen presentation and disease onset. The finding of this citrullinated vimentin peptide in the colonic tissue of RA patients, underlines the involvement of citrullination and vimentin in the RA etiology, as well as the systemic nature of the disease.

We successfully demonstrated the use of high throughput LC-MS technologies in the identification of statistically significantly abundant proteins in samples as complex as human colonic

tissue. A study including a larger cohort of patients, and restricting the analysis to e.g. the epithelial cells, will likely identify more significantly differentiating proteins between UC and the healthy state. The addition of CD patients in the study could also open for the possibility of identifying diagnostic markers for early and accurate differentiation of CD and UC patients. Also, a study implementing enrichment techniques for citrullinated peptides and proteins will likely identify more citrullination events.

3. Concluding Remarks & Outlook

The three peer reviewed publications and additional unpublished data included in this thesis demonstrate how recent development in proteomic strategies have allowed the identification of protein changes in samples as complex as human tissue. I have demonstrated how citrullination events can be reliably identified by including several identification criteria, including a tryptic missed cleavage, and demonstrates a potential pitfall of using the mentioned strategy. Furthermore, the strength of combining several complementary state of the art high throughput technologies, such as proteomic and transcriptomic, has been demonstrated. In my opinion, such combinations are necessary for eventually understanding systems as complex and composite as the human immune system.

The etiology of the inflammatory diseases remains poorly understood. As UC and RA share several genetic loci and known environmental triggers, I compared colonic biopsies from both groups to colonic biopsies from controls. The proteomic analysis of colonic biopsies from RA and UC patients revealed overall similarities in protein abundances between the biopsies from the RA patients and the controls. It has been hypothesized that the initial triggering of RA is not taking place in the joints but in the lungs. However, studies have suggested that the gut microbiota is involved in the initial triggering of the disease. Citrullinated vimentin was found with increased abundance in the colonic tissue of RA patients compared to the controls, which could indicate that the initial RA triggering is not limited to one location in the body, but can take place in many locations. The otherwise high degree of similarity found between the biopsies from the RA patients and the controls, is in contrast to the biopsies from the UC patients, where the mucosal immune system response to infections was found to be highly activated. Multiple proteins involved in NETs were found in higher abundance compared to in healthy mucosal tissue. The finding thereby support the current understanding in the literature, that IBD is caused by a complex interplay between the gut microbiota and the immune system, where an altered gut microbiota has been found during active IBD. However, what triggers the alteration from a commensal to pathogenic host-microbial relationship, remains to be understood. As mentioned, there is significantly cross-communication and co-modulation of the epithelial cells and the intestinal gut microbiota, and information on an altered composition of the gut microbiota alone does not yield information on changed host-microbial interactions. Therefore, the next logical step might be to investigate the host-microbial interactions in relation to IBD, which can be accomplished through the incorporation of metabolomic, transcriptomic and proteomic data. A

combined dataset would allow us to probe the host-microbial interactions, altered gut microbiota functions, and differentiating pathways, and would provide new insights into the pathogenesis of the IBD as well as other inflammatory diseases.

References

1. **Bennike, T.**, Birkelund, S., Stensballe, A. & Andersen, V. Biomarkers in inflammatory bowel diseases: Current status and proteomics identification strategies. *World J Gastroenterol* **20,** 3231–3244 (2014).
2. **Bennike, T.** *et al.* Optimizing the Identification of Citrullinated Peptides by Mass Spectrometry: Utilizing the Inability of Trypsin to Cleave after Citrullinated Amino Acids. *Journal of Proteomics & Bioinformatics* **06,** (2013).
3. **Bennike, T.** *et al.* A normative study of the synovial fluid proteome from healthy porcine knee joints. *Journal of Proteome Research* (2014).
4. Lettre, G. & Rioux, J. D. Autoimmune diseases: insights from genome-wide association studies. *Human molecular genetics* **17,** R116 (2008).
5. Bach, J.-F. The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases. *New England Journal of Medicine* **347,** 911–920 (2002).
6. Eaton, W. W., Rose, N. R., Kalaydjian, A., Pedersen, M. G. & Mortensen, P. B. Epidemiology of Autoimmune Diseases in Denmark. *J Autoimmun* **29,** 1–9 (2007).
7. Jussila, A., Virta, L. J., Pukkala, E. & Färkkilä, M. A. Mortality and causes of death in patients with inflammatory bowel disease: A nationwide register study in Finland. *Journal of Crohn's and Colitis* (2014). doi:10.1016/j.crohns.2014.02.015
8. Odes, S. *et al.* Cost Analysis and Cost Determinants in a European Inflammatory Bowel Disease Inception Cohort With 10 Years of Follow-up Evaluation. *Gastroenterology* **131,** 719–728 (2006).
9. Lévy, L., Fautrel, B., Barnetche, T. & Schaeverbeke, T. Incidence and risk of fatal myocardial infarction and stroke events in rheumatoid arthritis patients. A systematic review of the literature. *Clin. Exp. Rheumatol.* **26,** 673–679 (2008).
10. Lettre, G. & Rioux, J. D. Autoimmune diseases: insights from genome-wide association studies. *Human molecular genetics* **17,** R116 (2008).
11. Sarkar, S. & Fox, D. A. Targeting IL-17 and Th17 Cells in Rheumatoid Arthritis. *Rheumatic Disease Clinics of North America* **36,** 345–366 (2010).
12. Shen, W. & Durum, S. K. Synergy of IL-23 and Th17 Cytokines: New Light on Inflammatory Bowel Disease. *Neurochem Res* **35,** 940–946 (2010).
13. Rosenfeld, G. & Bressler, B. The truth about cigarette smoking and the risk of inflammatory bowel disease. *Am. J. Gastroenterol.* **107,** 1407–1408 (2012).
14. Wegner, N. *et al.* Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. *Immunological Reviews* **233,** 34–54 (2010).
15. Makrygiannakis, D. *et al.* Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. *Ann Rheum Dis* **67,** 1488–1492 (2008).
16. Klareskog, L., Malmström, V., Lundberg, K., Padyukov, L. & Alfredsson, L. Smoking, citrullination and genetic variability in the immunopathogenesis of rheumatoid arthritis. *Seminars in Immunology* **23,** 92–98 (2011).

17. Klareskog, L. *et al.* A new model for an etiology of rheumatoid arthritis: Smoking may trigger HLA–DR (shared epitope)–restricted immune reactions to autoantigens modified by citrullination. *Arthritis & Rheumatism* **54,** 38–46 (2006).
18. Klareskog, L., Rönnelid, J., Lundberg, K., Padyukov, L. & Alfredsson, L. Immunity to Citrullinated Proteins in Rheumatoid Arthritis. *Annual Review of Immunology* **26,** 651–675 (2008).
19. Khor, B., Gardet, A. & Xavier, R. J. Genetics and pathogenesis of inflammatory bowel disease. *Nature* **474,** 307–317 (2011).
20. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491,** 119–124 (2012).
21. Xavier, R. J. & Podolsky, D. K. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448,** 427–434 (2007).
22. Iskandar, H. N. & Ciorba, M. A. Biomarkers in Inflammatory Bowel Disease: Current Practices and Recent Advances. *Transl Res* **159,** 313–325 (2012).
23. Ehrentraut, S. F. & Colgan, S. P. Implications of protein post-translational modifications in IBD. *Inflammatory Bowel Diseases* **18,** 1378–1388 (2012).
24. Poulsen, N. *et al.* Comparative analysis of inflamed and non-inflamed colon biopsies reveals strong proteomic inflammation profile in patients with ulcerative colitis. *BMC Gastroenterology* **12,** 76 (2012).
25. Thompson, A. I. & Lees, C. W. Genetics of ulcerative colitis. *Inflammatory Bowel Diseases* **17,** 831–848 (2011).
26. Burisch, J., Jess, T., Martinato, M. & Lakatos, P. L. The burden of inflammatory bowel disease in Europe. *Journal of Crohn's and Colitis* **7,** 322–337 (2013).
27. Burisch, J., Jess, T., Martinato, M. & Lakatos, P. L. The burden of inflammatory bowel disease in Europe. *Journal of Crohn's and Colitis* **7,** 322–337 (2013).
28. Shivananda, S. *et al.* Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). *Gut* **39,** 690–697 (1996).
29. Molodecky, N. A. *et al.* Increasing Incidence and Prevalence of the Inflammatory Bowel Diseases With Time, Based on Systematic Review. *Gastroenterology* **142,** 46–54.e42 (2012).
30. Betteridge, J. D., Armbruster, S. P., Maydonovitch, C. & Veerappan, G. R. Inflammatory Bowel Disease Prevalence by Age, Gender, Race, and Geographic Location in the U.S. Military Health Care Population. *Inflammatory Bowel Diseases June 2013* **19,** 1421–1427 (2013).
31. Molodecky, N. A. *et al.* Increasing Incidence and Prevalence of the Inflammatory Bowel Diseases With Time, Based on Systematic Review. *Gastroenterology* **142,** 46–54.e42 (2012).
32. Lewis, J. D. The Utility of Biomarkers in the Diagnosis and Therapy of Inflammatory Bowel Disease. *Gastroenterology* **140,** 1817–1826.e2 (2011).
33. Vatn, M. H. Natural history and complications of IBD. *Curr Gastroenterol Rep* **11,** 481–487 (2009).
34. Meuwis, M.-A. *et al.* Biomarker discovery for inflammatory bowel disease, using proteomic serum profiling. *Biochemical Pharmacology* **73,** 1422–1433 (2007).

35. Han, N.-Y. *et al.* Label-free quantification for discovering novel biomarkers in the diagnosis and assessment of disease activity in inflammatory bowel disease. *Journal of Digestive Diseases* **14,** 166–174 (2013).
36. Vermeire, S., Van Assche, G. & Rutgeerts, P. Laboratory markers in IBD: useful, magic, or unnecessary toys? *Gut* **55,** 426–431 (2006).
37. Yau, Y., Leong, R. W., Zeng, M. & Wasinger, V. C. Proteomics and metabolomics in inflammatory bowel disease. *Journal of Gastroenterology and Hepatology* **28,** 1076–1086 (2013).
38. Berry, D. & Reinisch, W. Intestinal microbiota: A source of novel biomarkers in inflammatory bowel diseases? *Best Practice & Research Clinical Gastroenterology* **27,** 47–58 (2013).
39. Mendoza, J. L. & Abreu, M. T. Biological markers in inflammatory bowel disease: Practical consideration for clinicians. *Gastroentérologie Clinique et Biologique* **33, Supplement 3,** S158–S173 (2009).
40. Han, N.-Y., Kim, E. H., Choi, J., Lee, H. & Hahm, K.-B. Quantitative proteomic approaches in biomarker discovery of inflammatory bowel disease. *Journal of Digestive Diseases* **13,** 497–503 (2012).
41. Sandborn, W. J. *et al.* Evaluation of serologic disease markers in a population-based cohort of patients with ulcerative colitis and Crohn's disease. *Inflammatory Bowel Diseases* **7,** 192–201 (2001).
42. Reumaux, D. *et al.* Serological markers in inflammatory bowel diseases. *Best Practice & Research Clinical Gastroenterology* **17,** 19–35 (2003).
43. Peyrin-Biroulet, L., Standaert-Vitse, A., Branche, J. & Chamaillard, M. IBD serological panels: facts and perspectives. *Inflamm. Bowel Dis.* **13,** 1561–1566 (2007).
44. Bernstein, C. N. *et al.* Assessing inflammatory bowel disease-associated antibodies in Caucasian and First Nations cohorts. *Can J Gastroenterol* **25,** 269–273 (2011).
45. Peeters, M. *et al.* Diagnostic value of anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease. *Am J Gastroenterol* **96,** 730–734 (2001).
46. Saverymuttu, S. H., Hodgson, H. J., Chadwick, V. S. & Pepys, M. B. Differing acute phase responses in Crohn's disease and ulcerative colitis. *Gut* **27,** 809–813 (1986).
47. Solem, C. A. *et al.* Correlation of C-reactive protein with clinical, endoscopic, histologic, and radiographic activity in inflammatory bowel disease. *Inflammatory Bowel Diseases* **11,** 707–712 (2005).
48. Burri, E. & Beglinger, C. Faecal calprotectin in the diagnosis of inflammatory bowel disease. *Biochemia Medica* 245–253 (2011). doi:10.11613/BM.2011.034
49. Schoepfer, A. M., Trummler, M., Seeholzer, P., Seibold-Schmid, B. & Seibold, F. Discriminating IBD from IBS: Comparison of the test performance of fecal markers, blood leukocytes, CRP, and IBD antibodies. *Inflammatory Bowel Diseases* **14,** 32–39 (2008).
50. Van Venrooij, W. J., van Beers, J. J. B. C. & Pruijn, G. J. M. Anti-CCP antibodies: the past, the present and the future. *Nature Reviews Rheumatology* **7,** 391–398 (2011).
51. Balakrishnan, L. *et al.* Differential proteomic analysis of synovial fluid from rheumatoid arthritis and osteoarthritis patients. *Clinical Proteomics* **11,** 1 (2014).

52. Hui, A. Y., McCarty, W. J., Masuda, K., Firestein, G. S. & Sah, R. L. A systems biology approach to synovial joint lubrication in health, injury, and disease. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **4,** 15–37 (2012).
53. Symmons, D. *et al.* The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. *Rheumatology* **41,** 793–800 (2002).
54. Widdifield, J. *et al.* The rising burden of rheumatoid arthritis surpasses rheumatology supply in Ontario. *Can J Public Health* **104,** e450–455 (2013).
55. Ozminkowski, R. J., Burton, W. N., Goetzel, R. Z., Maclean, R. & Wang, S. The Impact of Rheumatoid Arthritis on Medical Expenditures, Absenteeism, and Short-Term Disability Benefits: *Journal of Occupational and Environmental Medicine* **48,** 135–148 (2006).
56. Hermansson, M. *et al.* MS analysis of rheumatoid arthritic synovial tissue identifies specific citrullination sites on fibrinogen. *Proteomics Clin Appl* **4,** 511–518 (2010).
57. Van Venrooij, W. J., van Beers, J. J. B. C. & Pruijn, G. J. M. Anti-CCP antibodies: the past, the present and the future. *Nature Reviews Rheumatology* **7,** 391–398 (2011).
58. Van Beers, J. J. B. C. *et al.* The rheumatoid arthritis synovial fluid citrullinome reveals novel citrullinated epitopes in apolipoprotein E, myeloid nuclear differentiation antigen, and β-actin. *Arthritis & Rheumatism* **65,** 69–80 (2013).
59. Venrooij, W. J. & Zendman, A. J. W. Anti-CCP2 Antibodies: An Overview and Perspective of the Diagnostic Abilities of this Serological Marker for Early Rheumatoid Arthritis. *Clinical Reviews in Allergy & Immunology* **34,** 36–39 (2007).
60. Vossenaar, E. Anti-CCP antibodies, a highly specific marker for (early) rheumatoid arthritis. *Clinical and Applied Immunology Reviews* **4,** 239–262 (2004).
61. Scott, D. L., Wolfe, F. & Huizinga, T. W. Rheumatoid arthritis. *The Lancet* **376,** 1094–1108 (2010).
62. Sande, M. G. H. van de *et al.* Different stages of rheumatoid arthritis: features of the synovium in the preclinical phase. *Ann Rheum Dis* **70,** 772–777 (2011).
63. Ytterberg, A. J. *et al.* Shared immunological targets in the lungs and joints of patients with rheumatoid arthritis: identification and validation. *Annals of the Rheumatic Diseases* (2014). doi:10.1136/annrheumdis-2013-204912
64. Consortium, T. E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
65. Feng, J. T., Shang, S. & Beretta, L. Proteomics for the early detection and treatment of hepatocellular carcinoma. *Oncogene* **25,** 3810–3817 (2006).
66. Satsangl, J. *et al.* Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. *The Lancet* **347,** 1212–1217 (1996).
67. Nørregaard Jensen, O. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current Opinion in Chemical Biology* **8,** 33–41 (2004).
68. Wu, L. *et al.* Variation and genetic control of protein abundance in humans. *Nature* **499,** 79–82 (2013).
69. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat Biotech* **21,** 255–261 (2003).

70. Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509,** 575–581 (2014).
71. Honoré, B., Østergaard, M. & Vorum, H. Functional genomics studied by proteomics. *Bioessays* **26,** 901–915 (2004).
72. Hugot, J.-P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411,** 599–603 (2001).
73. Satsangi, J. *et al.* Two stage genome–wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* **14,** 199–202 (1996).
74. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).
75. Wasinger, V. C. *et al.* Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis* **16,** 1090–1094 (1995).
76. Nagaraj, N. *et al.* System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. *Mol Cell Proteomics* **11,** (2012).
77. Thakur, S. S. *et al.* Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation. *Mol Cell Proteomics* **10,** (2011).
78. Steen, H. & Mann, M. The abc's (and xyz's) of peptide sequencing. *Nat Rev Mol Cell Biol* **5,** 699–711 (2004).
79. Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology* **7,** (2011).
80. Lescuyer, P., Bianchi, L., Hochstrasser, D., Bini, L. & Sanchez, J.-C. Translational proteomics. *Journal of Proteomics* **75,** 4571–4572 (2012).
81. Gershon, P. D. Cleaved and Missed Sites for Trypsin, Lys-C, and Lys-N Can Be Predicted with High Confidence on the Basis of Sequence Context. *Journal of Proteome Research* **13,** 702–709 (2014).
82. Vandermarliere, E., Mueller, M. & Martens, L. Getting intimate with trypsin, the leading protease in proteomics: TRYPSIN IN PROTEOMICS. *Mass Spectrometry Reviews* 000–000 (2013). doi:10.1002/mas.21376
83. Kaletaş, B. K. *et al.* Sample preparation issues for tissue imaging by imaging MS. *Proteomics* **9,** 2622–2633 (2009).
84. Leon, I. R., Schwämmle, V., Jensen, O. N. & Sprenger, R. R. Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *Molecular & Cellular Proteomics* (2013). at <http://www.mcponline.org/content/early/2013/06/21/mcp.M112.025585.abstract>
85. Nielsen, M. L. *et al.* Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat Meth* **5,** 459–460 (2008).
86. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68,** 850–858 (1996).
87. Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protocols* **1,** 2856–2860 (2007).

88. Zhang, N. & Li, L. Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation. *Rapid Communications in Mass Spectrometry* **18,** 889–896 (2004).
89. Rabilloud, T. *et al.* Power and limitations of electrophoretic separations in proteomics strategies. *Mass Spectrom. Rev.* **28,** 816–843 (2009).
90. Havliš, J. & Shevchenko, A. Absolute Quantification of Proteins in Solutions and in Polyacrylamide Gels by Mass Spectrometry. *Anal. Chem.* **76,** 3029–3036 (2004).
91. Speicher, K., Kolbas, O., Harper, S. & Speicher, D. Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *J Biomol Tech* **11,** 74–86 (2000).
92. Klammer, A. A. & MacCoss, M. J. Effects of Modified Digestion Schemes on the Identification of Proteins from Complex Mixtures. *Journal of Proteome Research* **5,** 695–700 (2006).
93. Manza, L. L., Stamer, S. L., Ham, A.-J. L., Codreanu, S. G. & Liebler, D. C. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **5,** 1742–1745 (2005).
94. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Meth* **6,** 359–362 (2009).
95. Masuda, T., Tomita, M. & Ishihama, Y. Phase Transfer Surfactant-Aided Trypsin Digestion for Membrane Proteome Analysis. *Journal of Proteome Research* **7,** 731–740 (2008).
96. Masuda, T., Tomita, M. & Ishihama, Y. Phase Transfer Surfactant-Aided Trypsin Digestion for Membrane Proteome Analysis. *Journal of Proteome Research* **7,** 731–740 (2008).
97. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10,** 57–63 (2009).
98. Gupta, N. & Pevzner, P. A. False Discovery Rates of Protein Identifications: A Strike against the Two-Peptide Rule. *Journal of Proteome Research* **8,** 4173–4181 (2009).
99. Vaudel, M., Sickmann, A. & Martens, L. Current methods for global proteome identification. *Expert Review of Proteomics* **9,** 519–532 (2012).
100. De Ceuleneer, M., Van Steendam, K., Dhaenens, M. & Deforce, D. In vivo relevance of citrullinated proteins and the challenges in their detection. *PROTEOMICS* **12,** 752–760 (2012).
101. György, B., Tóth, E., Tarcsa, E., Falus, A. & Buzás, E. I. Citrullination: A posttranslational modification in health and disease. *The International Journal of Biochemistry & Cell Biology* **38,** 1662–1677 (2006).
102. Curis, E. *et al.* Almost all about citrulline in mammals. *Amino Acids* **29,** 177–205 (2005).
103. Hartline, D. K. & Colman, D. R. Rapid Conduction and the Evolution of Giant Axons and Myelinated Fibers. *Current Biology* **17,** R29–R35 (2007).
104. Sjöstrand, F. S. The lamellated structure of the nerve myelin sheath as revealed by high resolution electron microscopy. *Experientia* **9,** 68–69 (1953).
105. Kim, J. K. *et al.* Multiple sclerosis: an important role for post-translational modifications of myelin basic protein in pathogenesis. *Mol. Cell Proteomics* **2,** 453–462 (2003).

106. Krokhin, O. V., Antonovici, M., Ens, W., Wilkins, J. A. & Standing, K. G. Deamidation of -Asn-Gly- Sequences during Sample Preparation for Proteomics: Consequences for MALDI and HPLC-MALDI Analysis. *Anal. Chem.* **78,** 6645–6650 (2006).
107. De Ceuleneer, M., Van Steendam, K., Dhaenens, M., Elewaut, D. & Deforce, D. Quantification of Citrullination by Means of Skewed Isotope Distribution Pattern. *J. Proteome Res.* **11,** 5245–5251 (2012).
108. Hao, G. *et al.* Neutral loss of isocyanic acid in peptide CID spectra: A novel diagnostic marker for mass spectrometric identification of protein citrullination. *J Am Soc Mass Spectrom* **20,** 723–727 (2009).
109. Creese, A. J., Grant, M. M., Chapple, I. L. C. & Cooper, H. J. On-line liquid chromatography neutral loss-triggered electron transfer dissociation mass spectrometry for the targeted analysis of citrullinated peptides. *Analytical Methods* **3,** 259 (2011).
110. Grolund, M.-M., Lehtonen, O.-P., Eerola, E. & Kero, P. Fecal Microflora in Healthy Infants Born by Different Methods of Delivery: Permanent Changes in Intestinal Flora After Cesarean Delivery. *Journal of Pediatric Gastroenterology* **28,** 19–25 (1999).
111. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
112. Manichanh, C., Borruel, N., Casellas, F. & Guarner, F. The gut microbiota in IBD. *Nature Reviews Gastroenterology and Hepatology* **9,** 599–608 (2012).
113. Tlaskalová-Hogenová, H. *et al.* Commensal bacteria (normal microflora), mucosal immunity and chronic inflammatory and autoimmune diseases. *Immunology Letters* **93,** 97–108 (2004).
114. Hooper, L. V. & Macpherson, A. J. Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat Rev Immunol* **10,** 159–169 (2010).
115. Flint, H. J., Scott, K. P., Louis, P. & Duncan, S. H. The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* **9,** 577–589 (2012).
116. Toivanen, A. & Toivanen, P. Reactive arthritis. *Current opinion in rheumatology* **12,** 300–305 (2000).
117. Ivanov, I. I. & Littman, D. R. Segmented filamentous bacteria take the stage. *Mucosal Immunol* (2010). at <http://dx.doi.org/10.1038/mi.2010.3>
118. Arif, S. *et al.* Peripheral and Islet Interleukin-17 Pathway Activation Characterizes Human Autoimmune Diabetes and Promotes Cytokine-Mediated ?-Cell Death. *Diabetes* **60,** 2112–2119 (2011).
119. Tokura, Y., Mori, T. & Hino, R. Psoriasis and other Th17-mediated skin diseases. *J. UOEH* **32,** 317–328 (2010).
120. Hu, Y., Shen, F., Crellin, N. K. & Ouyang, W. The IL-17 pathway as a major therapeutic target in autoimmune diseases. *Annals of the New York Academy of Sciences* **1217,** 60–76 (2011).
121. Wang, Y.-L. *et al.* Targeting pre-ligand assembly domain of TNFR1 ameliorates autoimmune diseases – An unrevealed role in downregulation of Th17 cells. *Journal of Autoimmunity* **37,** 160–170 (2011).
122. Gerasimidis, K. *et al.* Decline in Presumptively Protective Gut Bacterial Species and Metabolites Are Paradoxically Associated with Disease Improvement in Pediatric Crohn's

Disease During Enteral Nutrition. *Inflamm. Bowel Dis.* (2014). doi:10.1097/MIB.0000000000000023
123. Ivanov, I. I. *et al.* Specific microbiota direct the differentiation of Th17 cells in the mucosa of the small intestine. *Cell Host Microbe* **4,** 337–349 (2008).
124. Morgan, X. C. *et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* **13,** R79 (2012).
125. Willing, B. P. *et al.* A Pyrosequencing Study in Twins Shows That Gastrointestinal Microbial Profiles Vary With Inflammatory Bowel Disease Phenotypes. *Gastroenterology* **139,** 1844–1854.e1 (2010).
126. Gevers, D. *et al.* The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host & Microbe* **15,** 382–392 (2014).
127. Mukhopadhya, I., Hansen, R., El-Omar, E. M. & Hold, G. L. IBD-what role do Proteobacteria play? *Nat Rev Gastroenterol Hepatol* **9,** 219–230 (2012).
128. Berry, D. & Reinisch, W. Intestinal microbiota: A source of novel biomarkers in inflammatory bowel diseases? *Best Practice & Research Clinical Gastroenterology* **27,** 47–58 (2013).
129. Kostic, A. D., Xavier, R. J. & Gevers, D. The Microbiome in Inflammatory Bowel Disease: Current Status and the Future Ahead. *Gastroenterology* (2014). doi:10.1053/j.gastro.2014.02.009
130. Fava, F. & Danese, S. Intestinal microbiota in inflammatory bowel disease: Friend of foe? *World J Gastroenterol* **17,** 557–566 (2011).
131. Abraham, C. & Medzhitov, R. Interactions Between the Host Innate Immune System and Microbes in Inflammatory Bowel Disease. *Gastroenterology* **140,** 1729–1737 (2011).
132. Turner, J. R. Molecular Basis of Epithelial Barrier Regulation. *Am J Pathol* **169,** 1901–1909 (2006).
133. RUSSO, J. M. *et al.* Distinct Temporal-Spatial Roles for Rho Kinase and Myosin Light Chain Kinase in Epithelial Purse-String Wound Closure. *Gastroenterology* **128,** 987–1001 (2005).
134. R, M., S, C. & Jl, M. Rapid barrier restitution in an in vitro model of intestinal epithelial injury. *Lab Invest* **60,** 237–244 (1989).
135. Artis, D. Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nat Rev Immunol* **8,** 411–420 (2008).
136. Vaishnava, S., Behrendt, C. L., Ismail, A. S., Eckmann, L. & Hooper, L. V. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proc Natl Acad Sci U S A* **105,** 20858–20863 (2008).
137. Pastorelli, L., De Salvo, C., Mercado, J. R., Vecchi, M. & Pizarro, T. T. Central Role of the Gut Epithelial Barrier in the Pathogenesis of Chronic Intestinal Inflammation: Lessons Learned from Animal Models and Human Genetics. *Front Immunol* **4,** (2013).
138. Mantovani, A., Cassatella, M. A., Costantini, C. & Jaillon, S. Neutrophils in the activation and regulation of innate and adaptive immunity. *Nat Rev Immunol* **11,** 519–531 (2011).
139. Brinkmann, V. *et al.* Neutrophil Extracellular Traps Kill Bacteria. *Science* **303,** 1532–1535 (2004).

140. Urban, C. F. *et al.* Neutrophil Extracellular Traps Contain Calprotectin, a Cytosolic Protein Complex Involved in Host Defense against Candida albicans. *PLoS Pathogens* **5,** e1000639 (2009).
141. O'Donoghue, A. J. *et al.* Global Substrate Profiling of Proteases in Human Neutrophil Extracellular Traps Reveals Consensus Motif Predominantly Contributed by Elastase. *PLoS ONE* **8,** e75141 (2013).
142. Atkinson, A. J. *et al.* Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther* **69,** 89–95 (2001).
143. Leon, I. R., Schwämmle, V., Jensen, O. N. & Sprenger, R. R. Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *Molecular & Cellular Proteomics* (2013). at <http://www.mcponline.org/content/early/2013/06/21/mcp.M112.025585.abstract>
144. Urbaniak, G. C. & Plous, S. Research Randomizer (Version 4.0) [Computer software]. (2013). at <http://www.randomizer.org/>
145. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).
146. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98,** 5116–5121 (2001).
147. Jensen, L. J. *et al.* STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37,** D412–D416 (2009).
148. Bhatia, V. N., Perlman, D. H., Costello, C. E. & McComb, M. E. Software tool for researching annotations of proteins: open-source protein annotation software with data visualization. *Anal. Chem.* **81,** 9819–9823 (2009).
149. Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9,** 488 (2008).
150. J. C., Oliveros. VENNY. An interactive tool for comparing lists with Venn Diagrams. (2007). at <http://bioinfogp.cnb.csic.es/tools/venny/>
151. Urbaniak, M. D., Martin, D. M. A. & Ferguson, M. A. J. Global Quantitative SILAC Phosphoproteomics Reveals Differential Phosphorylation Is Widespread between the Procyclic and Bloodstream Form Lifecycle Stages of Trypanosoma brucei. *J Proteome Res* **12,** 2233–2244 (2013).
152. Sylvestersen, K. B., Horn, H., Jungmichel, S., Jensen, L. J. & Nielsen, M. L. Proteomic analysis of arginine methylation sites in human cells reveals dynamic regulation during transcriptional arrest. *Molecular & Cellular Proteomics* mcp–O113 (2014).
153. Dastidar, E. G. *et al.* Comprehensive Histone Phosphorylation Analysis and Identification of Pf14-3-3 Protein as a Histone H3 Phosphorylation Reader in Malaria Parasites. *PLoS ONE* **8,** e53179 (2013).
154. Lundby, A. *et al.* Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. *Nat Commun* **3,** 876 (2012).
155. Spitzer, M., Wildenhain, J., Rappsilber, J. & Tyers, M. BoxPlotR: a web tool for generation of box plots. *Nat Meth* **11,** 121–122 (2014).

156. Post, S. van der & Hansson, G. C. Membrane Protein Profiling of Human Colon Reveals Distinct Regional Differences. *Mol Cell Proteomics* mcp.M114.040204 (2014). doi:10.1074/mcp.M114.040204
157. Khor, T. S., Fujita, H., Nagata, K., Shimizu, M. & Lauwers, G. Y. Biopsy interpretation of colonic biopsies when inflammatory bowel disease is excluded. *Journal of Gastroenterology* **47,** 226–248 (2012).
158. Geboes, K. Histopathology of Crohn's disease and ulcerative colitis. *Inflammatory Bowel Disease. Edinburgh, London, Melbourne: Churchill Livingstone* 255–276 (2003).
159. Han, N.-Y. *et al.* Label-free quantification for discovering novel biomarkers in the diagnosis and assessment of disease activity in inflammatory bowel disease. *Journal of Digestive Diseases* **14,** 166–174 (2013).
160. Leach, S. T. *et al.* Serum and mucosal S100 proteins, calprotectin (S100A8/S100A9) and S100A12, are elevated at diagnosis in children with inflammatory bowel disease. *Scand J Gastroenterol* **42,** 1321–1331 (2007).
161. Barceló-Batllori, S. *et al.* Proteomic analysis of cytokine induced proteins in human intestinal epithelial cells: Implications for inflammatory bowel diseases. *PROTEOMICS* **2,** 551–560 (2002).
162. Hsieh, S.-Y. *et al.* Comparative proteomic studies on the pathogenesis of human ulcerative colitis. *PROTEOMICS* **6,** 5322–5331 (2006).
163. Barceló-Batllori, S. *et al.* Proteomic analysis of cytokine induced proteins in human intestinal epithelial cells: Implications for inflammatory bowel diseases. *PROTEOMICS* **2,** 551–560 (2002).
164. Kane, S. V. *et al.* Fecal lactoferrin is a sensitive and specific marker in identifying intestinal inflammation. *The American Journal of Gastroenterology* **98,** 1309–1314 (2003).
165. Angriman, I. *et al.* Enzymes in feces: Useful markers of chronic inflammatory bowel disease. *Clinica Chimica Acta* **381,** 63–68 (2007).
166. Bang, H. *et al.* Mutation and citrullination modifies vimentin to a novel autoantigen for rheumatoid arthritis. *Arthritis & Rheumatism* **56,** 2503–2511 (2007).
167. Hao, G. *et al.* Neutral loss of isocyanic acid in peptide CID spectra: A novel diagnostic marker for mass spectrometric identification of protein citrullination. *J Am Soc Mass Spectrom* **20,** 723–727 (2009).
168. Tutturen, A. E. V. *et al.* A technique for the specific enrichment of citrulline-containing peptides. *Analytical Biochemistry* **403,** 43–51 (2010).
169. Marlies De Ceuleneer *et al.* Modification of citrulline residues with 2,3-butanedione facilitates their detection by liquid chromatography/mass spectrometry. *Rapid Communications in Mass Spectrometry* **25,** 1536–1542 (2011).
170. Tutturen, A. E. V., Holm, A. & Fleckenstein, B. Specific biotinylation and sensitive enrichment of citrullinated peptides. *Anal Bioanal Chem* **405,** 9321–9331 (2013).
171. Knudsen, A. D., **Bennike, T.**, Birkelund, S., Kjeldal, H., Otzen, D. E., Stensballe, A . Condenser: A statistical aggregation tool for multi-sample quantitative proteomic data from Matrix Science Mascot Distiller™. *Journal of Proteomics* **103,** 261–266 (2014).

172.Carlsen, T. G., **Bennike, T.**, Christiansen, G. & Birkelund, S. A role for anti-HSP60 antibodies in arthritis: a critical. *OA Arthritis* **1,** (2013).

173 Haslauer, C. M., Ayturk, U. M., **Bennike, T.**, Birkelund, S., Stensballe, A., Steen, H., Warman, M. L., Murray, M. M.,. Time-Course After ACL Injury: Proteomic Analysis of Synovial Fluid and RNASeq Analysis of Synovium to Identify Effects on Wound Healing and Osteoarthritis Development. *In preparation* (2014).

174.Bennike, T., Carlsen, T. G., Ellingsen, T. J., Bonderup, O. K., Glerup, H., Sennels, L., Bøgsted, M., Andersen, V., Birkelund, S., Stensballe, A. The intestinal mucosal proteome of ulcerative colitis is different from healthy controls and rheumatoid arthritis. *In preparation* (2014).

Appendices

A. Additional Scientific Work

A.1 Published Papers not Included in the PhD thesis

[171] Knudsen, A. D., **Bennike, T.**, Birkelund, S., Kjeldal, H., Otzen, D. E., Stensballe, A. Condenser: A statistical aggregation tool for multi-sample quantitative proteomic data from Matrix Science Mascot DistillerTM. Journal of Proteomics 103, 261–266 (2014).

[172] Carlsen, T. G., **Bennike, T.**, Christiansen, G. & Birkelund, S. A role for anti-HSP60 antibodies in arthritis: a critical. OA Arthritis 1, (2013).

[173] Haslauer, C. M., Ayturk, U. M., **Bennike, T.**, Birkelund, S., Stensballe, A., Steen, H., Warman, M. L., Murray, M. M., Time-Course After ACL Injury: Proteomic Analysis of Synovial Fluid and RNASeq Analysis of Synovium to Identify Effects on Wound Healing and Osteoarthritis Development (in preparation, 2014)

[174] **Bennike, T.**, Carlsen, T. G., Ellingsen, T. J., Bonderup, O. K., Glerup, H., Sennels, L., Bøgsted, M., Andersen, V., Birkelund, S., Stensballe, A. The intestinal mucosal proteome of ulcerative colitis is different from healthy controls and rheumatoid arthritis, (in preparation, 2014)

A.2 Talks

2013, lecture at Harvard Medical School, Children's Hospital Boston, USA: Citrullinated Proteins: Disease Relevance and issues in Mass Spectrometry Identification.

2013, lecture at Åben Forskerdag, Frederecia: Analyse af Tarmbiopsier hos Patienter med Kronisk Betændelse Lokaliseret til Tarm og Led.

2013, symposium lecture at the Danish Proteomics Society, University of Southern Denmark: Synovial Fluid Proteome Analysis and the Origin of Synovial Proteins.

2013, lecture at Aalborg University: Inflammatory Diseases – An Analysis of the Underlying Biological Triggers.

A.3 Posters Presented at Conferences and Symposiums

2014, European Crohn's and Colitis Organisation (ECCO), Copenhagen, Denmark: Inflammatory joint- and bowel diseases : a clinical proteomics study seeking to identify the underlying biological triggers.

2013, HUPO 12th Annual World Congress, Yokohama Japan: Impaired Tryptic Proteolytic Activity at Citrullinated Amino Acids

2013, Symposium Inflammation at Interfaces, Hamburg Germany: Specific Detection of Citrullinated Peptides on the Q Exactive MS using Neutral Loss Triggered Precursor Selection"

2012, 60th ASMS Conference on MS, Vancouver Canada: Analytical Methodologies for Identification and Characterisation of Melanotan II in illicit Market Samples

2012, HUPO 11th Annual World Congress, Boston USA: Specific Detection of Citrullinated Peptides on the Q-Exactive MS using Neutral Loss Triggered Precursor Selection

A.4 Funding and External Collaborators

The PhD is financed by the pharmaceutical company Ferring Denmark, Department of Biotechnology and Department of Health Science and Technology Aalborg University Denmark. The project is collaboration between Aalborg University, Copenhagen University, University of Southern Denmark, the Regional Hospitals Silkeborg and Viborg, and the pharmaceutical company Ferring. Finally, Dr. Hanno Steen at the Proteomics core at Children's Hospital Boston, Harvard Medical School is an external collaborator.

The following funds have supported different aspects of the project: The doctoral school Aalborg University, Otto Mønsteds Fond, Lundbeckfonden, Knud Højgaards Fond, Oticon, region midtjyllands sundhedsvidenskabelige forskningsfond, Svend Andersens fond, and inflammation at interfaces.

B. Papers Included in the PhD Thesis

World Journal of Gastroenterology

Online Submissions: http://www.wjgnet.com/esps/
bpgoffice@wjgnet.com
doi:10.3748/wjg.v20.i12.3231

World J Gastroenterol 2014 March 28; 20(12): 3231-3244
ISSN 1007-9327 (print) ISSN 2219-2840 (online)

TOPIC HIGHLIGHT

WJG 20th Anniversary Special Issues (3): Inflammatory bowel disease

Biomarkers in inflammatory bowel diseases: Current status and proteomics identification strategies

Tue Bennike, Svend Birkelund, Allan Stensballe, Vibeke Andersen

Tue Bennike, Svend Birkelund, Allan Stensballe, Department of Health Science and Technology, Aalborg University, 9220 Aalborg, Denmark
Vibeke Andersen, Institute of Regional Health Research, University of Southern Denmark, 5000 Odense C, Denmark
Vibeke Andersen, Organ Center, Hospital of Southern Jutland, 6200 Aabenraa, Denmark
Author contributions: Bennike T wrote and revised the article; Birkelund S, Stensballe A and Andersen V helped critical review and provided suggestions; Bennike T finalized the revision.
Correspondence to: Vibeke Andersen, MD, PhD, Organ Center, Hospital of Southern Jutland, Kresten Philipsens Vej 15, 6200 Aabenraa, Denmark. vibeke.andersen1@rsyd.dk
Telephone: +45-2980-1078 Fax: +45-8883-4488
Received: September 27, 2013 Revised: January 13, 2014
Accepted: February 20, 2014
Published online: March 28, 2014

Abstract

Unambiguous diagnosis of the two main forms of inflammatory bowel diseases (IBD): Ulcerative colitis (UC) and Crohn's disease (CD), represents a challenge in the early stages of the diseases. The diagnosis may be established several years after the debut of symptoms. Hence, protein biomarkers for early and accurate diagnostic could help clinicians improve treatment of the individual patients. Moreover, the biomarkers could aid physicians to predict disease courses and in this way, identify patients in need of intensive treatment. Patients with low risk of disease flares may avoid treatment with medications with the concomitant risk of adverse events. In addition, identification of disease and course specific biomarker profiles can be used to identify biological pathways involved in the disease development and treatment. Knowledge of disease mechanisms in general can lead to improved future development of preventive and treatment strategies. Thus, the clinical use of a panel of biomarkers represents a diagnostic and prognostic tool of potentially great value. The technological development in recent years within proteomic research (determination and quantification of the complete protein content) has made the discovery of novel biomarkers feasible. Several IBD-associated protein biomarkers are known, but none have been successfully implemented in daily use to distinguish CD and UC patients. The intestinal tissue remains an obvious place to search for novel biomarkers, which blood, urine or stool later can be screened for. When considering the protein complexity encountered in intestinal biopsy-samples and the recent development within the field of mass spectrometry driven quantitative proteomics, a more thorough and accurate biomarker discovery endeavor could today be performed than ever before. In this review, we report the current status of the proteomics IBD biomarkers and discuss various emerging proteomic strategies for identifying and characterizing novel biomarkers, as well as suggesting future targets for analysis.

Key words: Inflammatory bowel disease; Biomarker; Proteomics; Citrullination; Ulcerative colitis; Crohn's disease; Posttranslational modification

Core tip: Establishing the correct diagnose of Crohn's disease and ulcerative colitis (UC) patients remains troublesome, and correct and early medication is critical. No reliable biomarkes have been implemented in clinical usage, to distinguish between Crohn's disease patients and UC patients. Considering the protein complexity encountered in intestinal biopsy samples and the recent development within the field of quantitative proteomics, submitting the intestinal mucosa to a more thorough analysis has the potential to reveal new biomarkers.

Bennike T, Birkelund S, Stensballe A, Andersen V. Biomarkers in inflammatory bowel diseases: Current status and proteomics identification strategies. *World J Gastroenterol* 2014; 20(12): 3231-3244 Available from: URL: http://www.wjgnet.com/1007-9327/full/v20/i12/3231.htm DOI: http://dx.doi.org/10.3748/wjg.v20.i12.3231

INTRODUCTION

Inflammatory bowel diseases (IBD) are chronic gastrointestinal disorders. The two most common forms of IBD are Crohn's disease (CD) and ulcerative colitis (UC). Both disorders have great impact on the life quality of the affected individuals and for society, measured on lost labor and expenses to the health care system. Furthermore, new epidemiological data published in 2013 found that the incidence and prevalence of the diseases are still increasing[1]. The etiologies of CD and UC remain unclear, but involve a complex interplay between genetic and environmental factors[2-7]. The diagnosis can be delayed several years and may be difficult to make even for trained physicians, as no biomarkers or commercial tests capable of discriminating CD from UC patients have been implemented in clinical use[8-10]. Furthermore, an early and accurate diagnosis of IBD-patients is crucial, as *e.g.,* CD patients with extensive and deep ulcerations have a 5-fold higher risk of requiring colectomy compared to CD patients without extensive and deep ulcerations[11]. From 357 CD patients analyzed with computed tomography enterography, penetrating disease was found in 21% of the patients and extraintestinal manifestations in 19%[12,13]. Hence, there is a need for reliable and usable biomarkers for the early and better diagnosis and prognosis of the IBD diseases[4,8,14-18].

GENOMIC, TRANSCRIPTOMIC AND PROTEOMIC BIOMARKERS

In 2001, an NIH group defined a biomarker as "A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention."[19], usable for diagnostics, monitoring disease prognosis and disease monitoring and prediction. The human genome contains the code for the expressed gene products, including the proteins. Proteins function as the building blocks of the human cells and tissue, and are responsible for the majority of the biological functions[20]. Proteins, therefore, represent an obvious target for biomarker discovery studies. The human genome comprises approximately 20000 protein coding genes[21]. During protein synthesis, the DNA code is first transcribed into different RNA transcripts. Each gene can give rise to several RNA transcripts resulting in a total of roughly 100000 different RNA transcripts[22-24] (Figure 1), which in turn are translated into 100000 different proteins. After translation, most proteins are covalently modified at least once[25], and the final mature protein products are termed proteoforms. These so-called posttranslational modifications (PTMs) are often crucial to the correct physiological function of the given protein, and can determine activity state, localization, turnover and interaction with other proteins and substrates[23,25,26]. More than 200 distinct biologically relevant PTMs have been identified[27], so each RNA transcript can be more than 200 different proteoforms. The PTMs increases the complexity and diversity of the proteins tremendously (Figure 1). As a result, it is estimated that the human body contains more than one million different proteoforms[23], which constitutes the human proteome (all expressed proteins).

When searching for biomarkers, it is possible to analyze the target sample on the DNA level, the RNA transcript level or the protein level. Techniques for studying an organisms DNA code (genome) or RNA transcripts (transcriptome) have the advantage that entire genomes and transcriptomes can be sequenced and studied with great sensitivity, precision and coverage, and a number of biomarkers have been found for various diseases. Using genomic sequencing techniques, several CD and UC loci have been known for more than a decade, and the studies have greatly increased our knowledge of the IBDs[22,28,29]. Several cellular IBD-pathways have been identified, including pathways involved in barrier function, epithelial restitution, microbial defense, immune regulation, reactive oxygen species generation, autophagy, and finally various stress and metabolic pathways associated with cellular homeostasis, reviewed by Khor *et al*[3]. However, as mentioned no IBD biomarkes capable of differentiating CD from UC have been implemented in daily clinical usage, and the impact of the genomic studies on the treatment and diagnosis of the IBDs has been questioned[30-32].

Proteins represent an obvious target for biomarker discovery studies, and as PTMs dramatically increases the diversity and in many cases function of the mature proteins, they represent a promising area for IBD biomarker studies. PTMs are introduced after translation of the RNA transcripts (Figure 1), hence analyzing DNA and RNA transcripts does not directly provide information about the PTMs. A key technique capable of measuring absolute and relative protein quantification in complex protein mixtures in a high-throughput manner, as well as identify several PTMs, is bottom-up mass spectrometry (MS) based proteomics[24,33]. Proteomics is the large-scale identification of proteins, and can often cover the study of all expressed proteins by an organism (the proteome). The bottom-up MS strategy is based on measuring the mass-to-charge ratios (m/z) of peptides derived from proteins which have been enzymatically cleaved into minor peptides. From the measured m/z's the molecular weight of the intact peptides can be calculated[25]. In addition to calculating the intact masses, the peptides are collided with an inert gas which fragments the peptides, and the fragment m/z's are measured. The proteins in the sample are subsequently identified by searching the

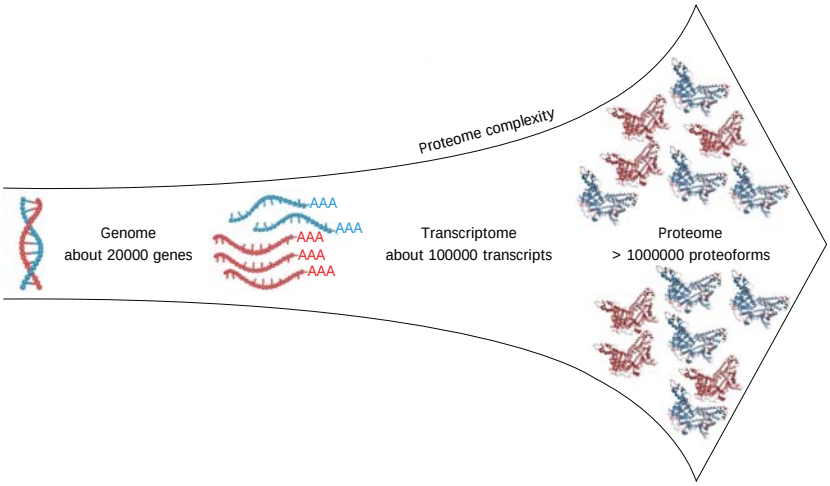


Figure 1 Major increase is encountered in the proteome complexity, from genes to RNA transcripts and finally to the mature, often posttranslational modification modified, proteins (proteoforms).

peptide masses and fragment m/z's against an *in silico* generated database, inferred from a reference database of protein sequences. By matching the *in silico* calculated peptide masses and fragment m/z's to the measured, the peptides and hence the proteins, are identified. For a more thorough description, we refer to the review by Steen *et al*[34]. The process can be performed in a quantitative manner to allow for relative or absolute quantitation of the proteins, using different strategies[34]. MS can in this way be used to identify proteins, as well as PTMs that change the molecular weight of the protein and can provide the amino acid position of the modification[25]. Previously, proteomics has been limited mainly by the speed and sensitivity of the mass spectrometers. However, recent development within the field of MS has allowed for the identification of nearly all expressed proteins of complex organisms, such as yeast, within a few hours of measuring time, identifying and quantifying several thousand proteins[33,35]. When considering the protein complexity encountered in the human intestinal tissue, an obvious place to search for biomarkers, and the recent development in the field of MS, a thorough analysis of PTMs and protein abundances in healthy and diseased state could be conducted. Biomarkers found in the intestine could then be searched for in more easily obtained sample material, such as blood or stool[6,10,31,36-39]. Antibodies to identified biomarkers for CD and UC found by proteomics can be generated for development of immunoassays and immunohistochemistry for evaluating the markers clinical use in routine tests less expensive than sequencing genomes, transcriptomes or MS driven proteomics.

This review reports known biomarkers for the IBDs, but will focus on the newly identified proteomics biomarkers and emerging proteomics strategies for identifying and characterizing novel IBD biomarkers.

DIAGNOSIS OF INFLAMMATORY BOWEL DISEASE AND KNOWN BIOMARKERS

Numerous biomarkers are known and used for the IBDs (Table 1); however, no single biomarker is able to diagnose IBD or to distinguish CD from UC patients with a high specificity and sensitivity[8-10,14]. CD is characterized by chronic inflammation in any part of the gastrointestinal tract. Most commonly the terminal ileum or the perianal region are inflamed, and in a non-continuous manner. Histologically, CD shows thickened submucosa, transmural inflammation, fissuring ulceration and non-caseating granulomas. UC, on the other hand, is characterized by inflammation limited to the colon, spreading continuously from the rectum and various distance proximal, and histology shows superficial inflammatory changes limited to the mucosa and submucosa with inflammation of crypts (cryptitis) and crypt abscesses[3]. There is currently no single "gold standard" diagnostic test or examination to differentiate CD and UC. Instead, diagnosis is based on a combination of symptoms, clinical examinations, laboratory findings, radiology, and endoscopy with histology, which also is used to assess severity and to predict the outcome of disease. Even when the tests are performed by expert clinicians they can result in diagnostic uncertainty[10,14,15,17,40]. This section will report some of the biomarkers commonly used to diagnose IBD. For a review of additional IBD biomarkers we refer to the work of Iskandar *et al*[4].

Antibodies and serum biomarkers

The two best-studied serological markers in IBD patients are anti-*Saccharomyces cerevisiae* antibodies (ASCA) and anti-neutrophil cytoplasmic antibody (ANCA)[41].

ASCA is an antibody with affinity for antigens in the

Table 1 Known common inflammatory bowel disease biomarkers

Biomarker	Specificity	Usability
Serum biomarkers		
ASCA	39%-79% of CD patients positive, 5%-15% UC patients[41-43]	14%-18% of controls tested positive, limiting the diagnostic value[44]
pANCA	20%-85% of UC patients positive, 2%-28% of the CD patients[41,42,45]	32% of controls tested positive, limiting the diagnostic value[44]
CRP	Marker for acute inflammation	Cannot differentiate CD from UC. However, usable for monitoring disease state[48-50]
Fecal biomarkers		
Calprotectin	Sensitive marker for intestinal inflammation[8,17,40]	Cannot differentiate CD from UC. Used to monitor disease state[17]
Lactoferrin	Can distinguish active IBD from inactive IBD and irritable bowel syndrome[60]	Unspecific for CD and UC. However, usable for monitoring disease state[60]

ASCA: Anti-*Saccharomyces cerevisiae* antibodies; ANCA: Anti-neutrophil cytoplasmic antibody; IBD: Inflammatory bowel diseases; UC: Ulcerative colitis; CD: Crohn's disease; CRP: C-reactive protein.

cell wall of the yeast *Saccharomyces cerevisiae*. In comparison to UC patients, CD patients are often positive for ASCA (Table 1)[41-43]. However, a substantial amount of healthy controls are also positive for ASCA positive[44], indicating that specificity and sensitivity for CD patients are relatively low; limiting the diagnostic value of the marker in differentiating CD from UC.

ANCAs are antibodies with affinity for neutrophil granules. The antibodies have been found in a variety of immune conditions, including Wegener's granulomatosis and rheumatoid arthritis (RA)[4]. When staining for ANCA, different patterns have been observed for UC and CD patients using immunofluorescence microscopy (Table 1), and mainly UC patients display perinuclear ANCA (pANCA) staining compared to CD patients[41,42,45]. Nonetheless, like the case of ASCA, a substantial amount of healthy controls are pANCA positive[44].

Lastly, C-reactive protein (CRP) is one of several proteins that increase in serum upon acute phase IBD. CRP is almost exclusively produced in the liver, upon stimulation by interleukin (IL)-6, tumor necrosis factor (TNF)-alpha and IL-1-beta produced at the site of inflammation. As such, an increased CRP-level is a marker for inflammation, but is not specific for CD or UC[8,40,46,47]. In some cases, but far from always, CD is associated with a strong CRP serum increase, whereas UC usually only results in a modest response. However, the difference insufficient to differentiate CD patients from UC patients[48-50], and the reason for the different responses remains to be thoroughly accounted for[40].

Other serum biomarkers used include white blood cell count, platelets, and albumin, which are all non-specific for IBD and can be seen in inflammatory diseases and cell stress[40]. More CD serologic markers are described in the review by Tamboli *et al*[51].

Fecal biomarkers

Stools are in direct contact with the inflamed intestinal area and site for the gut microbiome, both from which potential biomarkers are likely to originate. This is in contrast to serum biomarkers, which could increase on account of a variety of conditions, making stools an obvious place to search for biomarkers[40]. Fecal markers are especially useful for the diagnosis of CD patients, where the inflammation is patchy, may affect any part of the gastrointestinal tract, and therefore might be missed by colonoscopy[52]. The host-microbe interactions have been recognized as central for understanding human physiological diversity, and the human microbiome project has been launched to unravel the medical significance of the human microbiome[53]. Several studies have identified certain bacterial groups which are more abundant (*Enterobacteriaceae*, *Ruminococcus gnavus*, and *Desulfovibrio*) or less abundant (*Faecalibacterium prausnitzii*, *Lachnospiraceae*, *and Akkermansia*) in IBD[16], implicating that the host-microbe interaction might be involved, reviewed by Rosenstiel[54]. Novel biomarkers with high sensitivity and specificity may, therefore, be identified from stools.

The two most commonly used fecal markers for IBD screening are calprotectin and lactoferrin (Table 1)[8]. Calprotectin is a calcium- and zinc-binding protein occurring in large amounts in neutrophil granulocytes, where it accounts for 5% of the proteins. It is a very stable marker and is resistant to colonic bacterial degradation, and can be stored at room temperature for more than a week[55]. The concentration of fecal calprotectin is proportional to the neutrophil cell infiltrate in the bowel mucosa, and it is a very sensitive marker for intestinal inflammation[8,17,40]. However, calprotectin is not a specific marker for CD or UC, and increased levels can also be found with neoplasia, other forms of IBD, infections, and polyps[17], as well as with use of non-steroidal anti-inflammatory drugs, increasing age[56] and upper gastrointestinal disease, such as small bowel bacterial overgrowth[57].

Lactoferrin is an iron-binding glycoprotein expressed by activated neutrophils[58]. During inflammation, lactoferrin is released by the injured tissue and has been found to modulate inflammation and act in the defense against infections as a part of the innate immune system[59]. It is resistant to degradation and proteolysis, and unaffected by freeze thaw cycles, making it a useful biomarker[17]. As such, it is an ideal marker for intestinal inflammation. However, like calprotein it is unspecific for CD and UC, but can distinguish active IBD from inactive IBD and irritable bowel syndrome[60]. Several studies report similar performance of calprotectin and lactoferrin tests[6,60-64], and neither can be used to differentiate CD from UC with a high sensitivity and specificity.

To sum up, no reliable biomarkers exist usable as a single "gold standard". Therefore, to establish a diagnosis, histological examination of biopsies from the terminal ileum and colon is typically used in combination with

patient disease history and one or more of the above mentioned markers[17,65]. Hence, much effort is invested in analyzing the IBDs using various strategies, to identify usable biomarkers and explain the disease etiologies.

KNOWN PROTEOMICS BIOMARKERS FOR INFLAMMATORY BOWEL DISEASE

Proteomics studies can be performed in a discovery-based manner, where relative protein abundance levels between two or more samples are detected, and PTMs can be identified. Recent development of proteomics platforms has brought the technology to the point where several thousand proteins can be identified and (relatively) quantified in a single analysis or a subset by targeted approaches[6,10,31,36-39]. As inflammation takes place in the intestine, the gut-tissue represents an obvious place to look for novel biomarkers, which afterwards may be searched for in for example feces and blood and used as a disease marker. Several proteomic studies have successfully been aimed at identifying IBD biomarkers to investigate disease etiologies and aid in establishing the correct diagnose of UC and CD patients (Table 2). However, until now none of the identified biomarkers have been implemented in daily use[15].

The first group to publish a discovery-based proteomics study of the IBDs was Barceló-Batllori *et al*[66] in 2002. The aim of the study was to identify potential cytokine regulated proteins in colon epithelial cells isolated from IBD patients, which might be involved in the pathogenesis of IBDs. Human adenocarcinoma cells were *in vitro* exposed to known cytokines expressed in IBD, namely interferon-gamma, IL-1-beta and IL-6 (TNF-alpha was excluded as it is known to induce apoptosis in such cells). Using proteomics, the protein profiles of the cells were analyzed before and after exposure to the cytokines. All proteins from the cells were first separated using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). By staining all protein in the gels, different samples (gels) can be compared in terms of protein abundance based on the staining intensities, and differentiating protein spots can be visually identified. Spots of interest were cut from the gel with a knife and the proteins were enzymatically digested to specific peptides using the protease trypsin (in-gel digestion). The digestion of proteins is an essential step for protein identification, as no MS technique currently exist that can identify thousands of intact proteins in a complex sample in a high throughput manner. This is only possible when using digested proteins (peptides). The proteins were identified based on the peptides using MS, with the technique called matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) MS (Figure 2A). MALDI-TOF MS is a sensitive technique, but it involves placing a few drops of the sample on a plate which is left to dry prior to analysis. During analysis a laser is used to evaporate small spots from the dried droplet and ions in the produced gas are analyzed by MS. In the study, several cytokine regulated proteins were identified. Subsequently, human epithelial cells were isolated from UC patients and CD patients. Based on the findings, the samples were analyzed for the enzyme indoleamine-2,3-dioxygene using antibodies by western blotting. The group found an overabundance of the enzyme indoleamine-2,3-dioxygene in CD and UC compared to normal mucosa, hypothesizing an involvement of the Kynurenine pathway of tryptophan metabolism in the IBDs. Indoleamine-2,3-dioxygene activity has furthermore been found to be essential in dendritic cells to induce co-cultured T cell apoptosis[66].

When analyzing protein spots cut from gels the MALDI-TOF MS method is applicable, but the analysis of an entire 2D-PAGE gel is unfeasible, due to the commonly several thousand detectable spots. The technique is therefore less suitable for high-throughput identification of many thousand proteins. Therefore, when analyzing digested 2D-PAGE gels one usually only investigates changing protein spots and omits any information regarding non-changing protein spots. Information regarding non-changing proteins might prove equally important as changing proteins for studies seeking to describe disease etiologies. However, for biomarker studies 2D-PAGE strategies represent a feasible and proven way of identifying biomarker candidates. MALDI-TOF MS can also be conducted using intact proteins without prior enzymatic protein digestion. A variant of MALDI-TOF MS is to spot the protein mixture on a modified surface, to which the intact proteins bind and subsequently the intact masses of the proteins can be obtained by MS. This technique is called surface-enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF MS) (Figure 2A). However, when studying intact proteins using MALDI-TOF MS or SELDI-TOF MS, one usually does not obtain identification of the detected signals.

Electrospray ionization (ESI) remains the only MS technique for identifying and quantifying several thousands of proteins in a high-throughput manner (Figure 2B). ESI involves spraying the digested proteins directly into the MS. By incorporating liquid chromatography (LC) with columns prior to the ESI process, the peptides can be separated and sequentially eluted over several hours. This gives the MS systems enough time to analyze a large proportion of the eluted peptides which subsequently can be identified. In this way, large-scale proteomic studies can be performed in a high-throughput manner using ESI LC-MS. These studies yield (relative) quantitative information of thousands of identified proteins in a single experiment, and thus might provide better information for explaining disease etiologies. In 2004, Hardwidge *et al*[67] published such a study, which was the first large scale proteomic analysis of a human cellular response to a pathogen. Discovery-based proteomics was applied to investigate the protein profiles (cellular response) of human Caco-2 intestinal epithelia cells before and after infection with *E. coli*. The group did not work directly with IBD, but the results are applicable to the diseases,

Table 2 Proteomics biomarker candidate studies and main findings

Ref.	Sample	Analysis	Findings and perspectives
Barcelo-Batllori *et al*[66], 2002	*In vitro* colon epithelial cells and purified epithelial cells from UC and CD patients	2D-PAGE protein quantitation, and in-gel digestion and MALDI-TOF MS and Western blot protein identification	The enzyme indoleamine-2,3-dioxygene was more abundant in cells from CD and UC patients compared to normal mucosa. Tryptophan and arginine metabolism may play a role in the IBDs
Hardwidge *et al*[67], 2004	Human Caco-2 intestinal epithelia cells before and after infection with *E. coli*	ESI LC-MS protein identification and quantitation, Western blot verification	125 proteins more abundant and 139 proteins less abundant after infection, some related to innate immune responses. These proteins might be relevant to look for in future biomarker studies
Hsieh *et al*[68], 2006	Colonic biopsies from UC, nonspecific infectious colitis patients and controls	2D-PAGE protein quantitation, and in-gel digestion and MALDI-TOF MS protein identification	6 proteins were found to be more abundant in UC and 13 less abundant. The result indicates that mitochondrial dysfunction might be involved in UC the etiology. Four biomarker candidates were identified, however, they require validation
Shkoda *et al*[69], 2007	Intestinal tissue cells purified from patients suffering from CD, UC, and colon cancer	2D-PAGE protein quantitation, and in-gel digestion and MALDI-TOF MS and Western blot identification	Proteins associated with signal transduction, stress response and energy metabolism were differently abundant in inflamed and non-inflamed tissue. 32% of the differentially regulated proteins were involved in energy metabolism
Meuwis *et al*[10], 2007	Serum from UC and CD patients	SELDI-TOF MS m/z signal profiling, MALDI-TOF MS and Western blot protein identification	Successful in differentiating CD from UC patients with a sensitivity of 85% and a specificity of 95% from several m/z signals. Four biomarker candidates were identified, all known acute inflammatory markers, limiting the diagnostic value. However, the feasibility of serum biomarker studies was demonstrated
Nanni *et al*[71], 2007	Serum from UC, CD patients and healthy controls	Solid-phase bulk protein extraction, MALDI-TOF MS signal profiling	Able to separate the three groups with 97% prediction results. The signals were not identified, but the feasibility of serum biomarker studies was demonstrated
Meuwis *et al*[70], 2008	Serum from responding and non-responding CD patients to infliximab	SELDI-TOF MS signal profiling, MALDI-TOF MS, Western blot and ELISA protein identification	Able to predict responders with a sensitivity of 79% and a specificity of 80%. Increased amount of PF4 was associated with non-response to infliximab with MS but not ELISA, so usability of PF4 as a biomarker seems limited
Nanni *et al*[72], 2009	Intestinal epithelial cells from CD patients and healthy controls	1D-PAGE and in-gel digestion, ESI LC-MS protein identification and quantitation	Proteins more abundant in CD patients include several proteins involved in inflammation processes, and less abundant include Annexin A1, involved in the anti-inflammatory action. Follow-up research is required to assess the feasibility of the biomarker candidates
Hatsugai *et al*[73], 2010	Peripheral blood mononuclear cells from UC and CD patients, and healthy controls	2D-PAGE quantitation, and in-gel digestion and MALDI-TOF MS protein identification	Successfully discriminated UC from CD based on seven differently present proteins, all associated with inflammation oxidation/reduction, the cytoskeleton, endocytotic trafficking and transcription. The biomarker candidates require validation using a larger number of patients, but seems promising
M'Koma *et al*[74], 2011	Mucosal and submucosal layers of samples originating from CC and UC patients	MALDI-TOF MS m/z signal characterization, no protein identification	Five m/z signals were detected in the submucosal layer, which could separate the two groups with an accuracy of 75 percent. The signals needs to be identified, however, the disease groups can be separated on basis of the mucosal and submucosal profiles
Presley *et al*[75], 2012	Microbes and human proteins at the intestinal mucosal-luminal interface from CD and UC patients, and healthy controls	Oligonucleotide ribosomal RNA fingerprinting, SELDI-TOF MS and MALDI-TOF MS identification	35% of the detected bacterial phylotypes were present in different amounts in the diseases, indicating the involvement of host-microbe interactions in IBD. The microbiome might prove useful as a target for therapy
Han *et al*[14], 2013	Colonic tissue biopsies of Korean IBD patients	ESI LC-MS protein identification with label-free quantitation	27 potential biomarkers were identified for UC, 37 biomarkers for CD and 11 proteins commonly associated with IBD. Three novel biomarkers were identified for active CD: Bone marrow proteoglycan, L-plastin and proteasome activator subunit 1. The biomarker candidates require validation, but might prove feasible as new diagnostic and therapeutic targets
Seeley *et al*[76], 2013	Histological tissue layers from UC and CC patients	MALDI-TOF MS m/z signal characterization, no protein identification	114 different m/z signals were found to be different between the two groups. The signals remain unidentified
Gazouli *et al*[77], 2013	Serum samples from responding and non-responding CD patients to infliximab treatment	2D-PAGE quantitation, and in-gel digestion and MALDI-TOF MS protein identification	15 differently abundant proteins between responders and non-responders to infliximab were identified. The biomarker candidates require further validation

IBD: Inflammatory bowel diseases; UC: Ulcerative colitis; CD: Crohn's disease; MALDI-TOF MS: Matrix-assisted laser desorption-ionization time-of-flight mass spectrometry; 2D-PAGE: Two-dimensional polyacrylamide gel electrophoresis; ESI: Electrospray ionization; SELDI-TOF MS: Surface-enhanced laser desorption/ionization time of flight mass spectrometry; LC: Liquid chromatography.

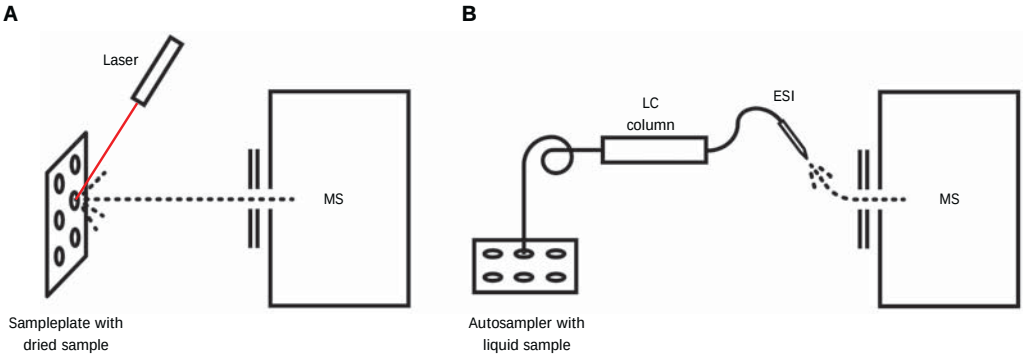


Figure 2 Two commonly used mass spectrometry techniques. A: Matrix-assisted laser desorption/surface-enhanced laser desorption/ionizatio time of flight mass spectrometry (MS), where the peptide or protein sample is dried on a target plate. Subsequently, a laser is used to evaporate the dried sample, and the generated gas phase ions are analyzed by the mass spectrometer; B: Liquid chromatography (LC)-electrospray ionization (ESI) MS, where the liquid peptide (or protein) sample is separated on a LC column, and sequentially eluted often over several hours. The eluted peptides are injected directly into the mass spectrometer by ESI and analyzed.

as the involvement of host-microbe interactions in the IBDs have been suggested[54]. The cells were lysed, and the lysates were chemically modified using chemical labels to allow for a relative comparison between the protein abundances measured by MS. Using ESI LC-MS, the group recorded 10921 peptide fragments mass spectra, from which they were able to identify 2000 proteins. Two hundred and sixty four proteins had a known biological function and were found to have at least a 2-fold abundance difference between infected and non-infected, roughly half were more abundant post infection. Some of the MS-findings were verified with western blots, and significant changes were found in amount of actin-related proteins before and after infection.

Even though ESI LC-MS has advantages in terms of high-throughput, many biomarker studies have successfully employed MALDI-TOF MS protein identification in IBDs. In 2006, Hsieh *et al*[68] applied discovery based proteomics using such a platform. The group analyzed the etiology and pathogenesis of UC using colonic biopsies to detect any significant difference in the protein profiles. The biopsies were obtained from four UC patients, three patients with nonspecific infectious colitis and five individuals with no obvious colonic disease. The proteins were separated by 2D-PAGE and a total of 1000 protein spots were compared visually between the diseased *vs* normal colon mucosa tissues. Forty protein-spots were found to be consistently different in intensity. Spots of interest were cut from the gel, tryptic digestion was performed and 19 proteins were identified using MALDI-TOF MS. Hereof, 13 identified proteins were less abundant in the UC-group and six proteins were more abundant. Eight of the less abundant proteins were identified as being mitochondrial proteins, suggesting that mitochondrial dysfunction might be involved in UC.

A year later in 2007, Shkoda *et al*[69] also identified a potential association between dysfunction in the energy metabolism and IBDs. The group applied a similar strategy and platform to investigate the loss of intestinal cell function, a critical component in the initiation and perturbation of chronic intestinal inflammation, and was the first to compare inflamed and non-inflamed tissue from the same patient. Intestinal cells were purified from intestinal tissue obtained from patients suffering from CD, UC, and colon cancer. The proteins were separated by 2D-PAGE and analyzed by MALDI-TOF MS and western blotting. 41 proteins were found to be differently abundant between inflamed and non-inflamed tissue, including proteins associated with signal transduction, stress response and energy metabolism. Thirty-two percent of all detected differentially regulated proteins associated with IBD were involved in energy metabolism. In 2007, Meuwis *et al*[10] published the first proteomic serum profiling study using SELDI-TOF MS in IBD, a variation of MALDI-TOF MS. The study included 30 patients with CD, 30 patients with UC, 30 inflammatory controls and 30 healthy controls. By characterizing the serum only by the m/z signals and not identified proteins with SELDI-TOF MS, the group was able to differentiate CD from UC with sensitivity of 85% (51/60) and specificity of 95% (57/60). Several of the unidentified signals were subsequently identified by MALDI-TOF MS, western blotting, and ELISA assay. Four biomarker candidates were identified: platelet aggregation factor 4 (PF4), myeloid related protein 8, fibrinopeptide A and haptoglobin alpha-2 subunit. All four proteins are known acute inflammatory markers to be expected in the IBDs, but the study succeeded in demonstrating that the separation of CD and UC patients based on serum markers is possible, highlighting the potential of serum profiling.

A year later, Meuwis *et al*[70] used the same platform and strategy to analyze if serum from 20 CD patients could be used to predict response to infliximab treatment. Infliximab is a monoclonal antibody against TNF-alpha, and was the first anti-TNF-alpha agent accepted for

IBD treatment. The protein profiles were characterized in serum prior to and post treatment with SELDI-TOF MS. The group verified the four previous biomarkers, and especially increased amount of PF4 was associated with non-response to infliximab. However, the association could not be confirmed by ELISA, and did not correlate significantly with other disease markers. Even so, the study was able to predict responders with a sensitivity of 79% (55/70) and a specificity of 80% (56/70). Even though the study did not succeed in identifying a usable biomarker for the prediction of responders, the study highlighted the potential in proteomic studies and response marker discovery.

In 2007, Nanni *et al*[71] optimized the methodological approach used to evaluate serum with MALDI-TOF MS. Using a solid-phase bulk protein extraction protocol followed by MALDI-TOF MS, they analyzed serum from 15 CD, 26 UC and 22 healthy individuals and were able to separate the three groups with 97% prediction results. Two years later, Nanni *et al*[72] conducted a study using high-throughput ESI LC-MS to investigate protein variations in the intestinal epithelial cells from CD patients. However, in contrast to Hardwidge *et al*[67] in 2004 who used chemical labelling of the peptides to measure the relative abundances, Nanni *et al*[72] employed a label-free strategy, and relied on the accurate detection of the peptide masses. In this way, significant savings can be achieved for large studies and the sample preparation protocols simplified. Intestinal epithelial cells were isolated from samples originating from two CD patients and two control patients. The cells were lysed and the proteins were separated by 1D-PAGE, where the proteins are separated only in one dimension in contrast to 2D-PAGE, which allowed the entire visualized gel lane to be cut into pieces and digested with trypsin. The resulting peptides were analyzed by ESI LC-MS and by comparing the peptide intensities, relative protein abundances could be calculated. Proteins which were found to be more abundant in the epithelial cells from CD patients include heat shock protein 70, tryptase alpha-1 precursor as well as several proteins involved in inflammation processes. The nuclear protein Annexin A1, involved in the anti-inflammatory action, and the malate dehydrogenase enzyme was found to be less abundant. The feasibility of the biomarker candidates remains to be validated. However, of great importance is the demonstration of the utility of label-free ESI LC-MS analysis for the identification of differences in protein abundances for IBD.

In 2010, Hatsugai *et al*[73] performed the first study which successfully discriminated UC from CD completely. The group analyzed peripheral blood mononuclear cells from 17 UC patients, 13 CD patients and 17 healthy controls. The proteins were separated by 2D-PAGE and more than 1000 protein spots were detected in each gel. Five hundred and forty-seven protein spots were selected for the quantitative analysis, and 34 protein spots were significantly different between the UC and CD groups. Using 58 protein spots, the UC and CD patients could be differentiated. The 58 protein spots were furthermore subjected to in-gel tryptic digestion followed by MALDI-TOF MS protein identification. Eleven of the proteins were successfully identified, and were found to be functionally related to inflammation, oxidation/reduction, the cytoskeleton, endocytotic trafficking and transcription. The profiles could, furthermore, predict disease severity and the UC patients' responses to treatment.

In 2011, M'Koma *et al*[74] analyzed mucosal and submucosal layers of samples originating from Crohn's colitis (CC) and UC, using MALDI-TOF MS. Five unknown m/z MS signals were detected, which could separate the two groups. The study did not identify the origin of the signals, but highlighted the possibility of finding biomarkers in the intestinal tissue.

As mentioned earlier, even though we are far from having a complete picture of the intestinal micro-biome, changes in the bacterial composition have been detected in IBD. In 2012, Presley *et al*[75] investigated the host-microbe interaction at the intestinal mucosal-luminal interface of 14 CD patients, 21 UC, and 16 healthy controls. The mucosa prevents microorganisms from entering the host tissue. Using a novel saline-lavage technique, saline was injected during colonscopy and extracted again to avoid interference from the intestinal layer contents resulting from a biopsy sample. The bacterial ribosomal RNA genes were analyzed by oligonucleotide fingerprinting and the proteins were analyzed by SELDI-TOF MS and MALDI-TOF MS. A combined proteome was constructed, constituting the proteomes from all detected organisms. Of the 3374 detected bacterial phylotypes, 35% significantly differentiated the diseases, indicating that host-microbe interactions might be involved in IBD, presenting new possibilities for diagnosis and therapy.

In 2013, Han *et al*[14] analyzed colonic tissue of Korean IBD patients in a high-throughput manner using ESI LC-MS and label-free quantitation. The study included four UC patients, three CD patients and two with inflammatory related polyps related to UC. The biopsies were homogenized and digested with trypsin without prior prefractionation and on average 324 proteins were identified for each group. Even though the number of identified proteins is relatively low considering the 2000 proteins Hardwidge *et al*[67] identified in 2004, 27 potential biomarkers were identified for UC, 37 biomarkers for CD and finally 11 proteins that were commonly associated with IBD. Three novel proteins, bone marrow proteoglycan, L-plastin and proteasome activator subunit 1 were identified as potential biomarkers for active CD. These biomarkes need validation, however, the feasibility of conducting high-throughput proteomics with label-free strategies in biomarker discovery was demonstrated.

A study published in 2013 by Seeley *et al*[76] investigated histological layers of 62 confirmed UC and CC tissues by MALDI-TOF MS. A total of 114 m/z MS signals were found to be statistically different between the two groups, however the signals have yet to be identified.

Finally, in 2013, Gazouli *et al*[77] published a study

where the response of 18 CD patients to infliximab treatment was correlated with known serum biomarkers. Serum samples were analyzed using 2D-PAGE, and 240 protein spot were selected for in-gel digestion and subsequent MALDI-TOF MS protein identification. The group was successful in identifying 15 proteins which were differentially present in the serum of CD patients depending on the response to infliximab. The proteins apolipoprotein A-I, apolipoprotein E, basic complement C4, plasminogen, serotransferrin, beta-2-glycoprotein 1, and clusterin were found to be more abundant in the patient groups with clinical and serological non-responders and responders, than in the group of patients with clinical and serological remission. Additionally, leucine-rich alpha-2-glycoprotein, vitamin D-binding protein, alpha-1B-glycoprotein and complement C1r subcomponent were found to be more abundant in the serum of the group of patients with remission. Interestingly, the group was unable to confirm the findings by Meuwis *et al*[70], that PF4 could be a biomarker for infliximab response, emphasizing that the biomarker candidates need further validation. Nonetheless, the study was successful in demonstrating the feasibility of identifying biomarkers in the serum usable to predict treatment outcome.

As apparent, many studies have successfully applied proteomic strategies to identify biomarkers, investigate IBD pathogenesis and identify prognostic markers in serum, stools, and tissue. Several biomarkers have been found (Table 2), most related to unspecific inflammation, and all biomarker candidates identified so far lacks follow-up validation studies. However, even though many of the identified biomarkers are related to inflammation, the studies have demonstrated the feasibility and potential of the proteomics platform in IBD, and given clues to the mechanisms of the IBDs. A few studies have successfully differentiated CD patients from UC patients. However, only based on unidentified m/z signals and not using identified protein or peptide biomarkers, from which the disease etiologies might be better explained. Nonetheless, these studies demonstrate the presence of usable biomarkers yet to be identified. Identified biomarkers hold the potential for designing diagnostic ELISA tests and protein array chips, where antibodies are used to detect the abundance of one or more antigens[78,79]. Such arrays could constitute new clinical tools for diagnosis, prognosis and identify novel targets for therapy.

The studies have demonstrated the presence of biomarkers, in serum, in the intestinal tissue and in stools. Many studies have aimed at performing global discovery-based proteomics in the intestinal tissue, and it has been demonstrated that high-throughput techniques such as ESI LC-MS, employing labelling or label-free quantitation are feasible ways to identify biomarkers in highly complex samples. The advantage of high-throughput protein identification and quantification strategies are especially apparent when disease etiologies are to be examined.

Furthermore, few studies have investigated the possible association between various PTMs and the IBD disease etiologies. Such an association is known from other inflammatory diseases; an example being the inflammatory joint-disease rheumatoid arthritis (RA) where the PTM citrullination is known to be involved in the etiology[80-83].

POSTTRANSLATIONAL MODIFICATIONS AS BIOMARKERS

Today, more than 200 distinct PTM's are known[84]. The PTMs are, to a large extend, important for the physiological function of the protein and the half-life of PTMs range from milliseconds to years[85]. Unfortunately, they are also often low abundance, highly diverse and complex, and thus can be challenging to detect and characterize[25,27,86]. Hence, PTMs represent promising targets for biomarker discovery studies. For a review on protein regulation by PTMs in the IBDs, we refer to the work by Ehrentraut *et al*[5]. Common *in vivo* PTMs include phosphorylation, which is a reversible modification of the amino acids tyrosine, serine and threonine. Phosphorylation is known to be involved in activation and inactivation of enzyme activity, modulation of molecular interactions and cell signaling through specific domains. Acetylation can target any N-terminal, and it is believed that 84% of all human proteins undergo this modification[87]. The PTM affects the protein stability, and histone acetylation is known to play a role in gene regulation. Glycosylation is another central PTM. It is reversible and known to be involved in cell-cell recognition and signaling, and regulation of proteins. Disulfide bond formation between two cysteines is a key element in the stabilization of proteins and protein complexes, such as, antibodies by forming intra- and intermolecular crosslinks. Deamidation of asparginine or glutamine is a possible regulator of protein-ligand and protein-protein interactions, and ubiquitination is a marker for protein recycling/destruction[25]. Several PTMs are known to be involved in the inflammatory responses, and PTMs could be involved in the IBD disease etiologies. Lastly, citrullination is the irreversible deimination of arginine into citrulline, *in vivo* catalyzed by the peptidylarginine deiminases, a calcium binding family of enzymes[88,89]. The exact role of the modification remains largely unknown, but the modification is believed to alter the fold of the proteins, change the protein polarity, and/or lead to denaturation in order to render the protein more prone to enzymatic degradation[80,88,89]. Citrullination has been associated with several diseases, including Alzheimer's disease[90], and RA where an anti-citrullinated protein antibody was identified[80-83]. Smoking has been associated with increased citrullination, and smoking is the best known environmental factor for the development of RA[91-95]. Several studies have, furthermore, associated smoking with an increased risk of developing CD and UC[96-100]. In RA, it is believed that citrullination of proteins results in the generation of new antigens being presented to the immune system, which in turn triggers an autoimmune response[83]. It therefore seems plausible that citrullination may have a similar role

in the IBDs as well as other inflammatory diseases. However, as with many PTMs the MS-driven detection of citrullinated proteins in a high-throughput manner is not straight forward[84,101-104]. Nonetheless, if disease-specific citrullinated proteins could be identified, these could be utilized in ELISA or protein array chips for prognostics and/or diagnostics. An example of the utilization of a similar biomarker is the diagnosis of RA patients, where the presence of anti-citrullinated protein antibodies in the serum is used to detect the disease with a sensitivity of 71% and specificity of 95%[80-83].

CONCLUSION

The diagnosis of UC and CD patients remains difficult, especially in the early stages of the diseases, and early and accurate diagnosis of IBD-patients is crucial. Several studies have successfully identified promising biomarkers in stools, serum and tissue, demonstrating the presence of IBD biomarkers. However, none of the identified biomarkers have been implemented in clinical daily use, and the diagnosis is based on a combination of disease history, colonscopy inflammation biomarkers and histological evaluation.

Few studies have aimed at investigating the global proteome of intestinal tissue using high-throughput techniques such as ESI LC-MS, and the potential of such analysis seems immense. The recent development within the field of high-throughput protein identification using MS, now allows for identifying and quantifying several thousand proteins in a few hours of analysis time. Besides protein abundances, PTMs represent promising targets for biomarker discovery studies. An analysis of tissue, serum or stools therefore seems promising to identify novel biomarkers. Such information could be used to make accurate diagnostic and prognostic tools to differentiate patient groups and predict treatment responses. Antibodies against one or more identified diagnostic targets could be used in ELISA or protein array chips, which in turn can be used to detect the abundance of the given antigen. Besides aiding physicians in making a correct diagnosis and treatment strategy, knowledge of disease specific proteins and PTMs might identify disease pathways and new targets for therapeutic agents, leading to improved pharmaceutical drugs.

Conclusively, protein identification and quantification using mass spectrometry holds great promise for the identification of novel diagnostic and prognostic biomarkers for the IBDs, and might help explain the disease etiologies, ultimately leading to improved treatment strategies.

ACKNOWLEDGMENTS

The authors would like to thank Lau Sennels, Gunna Christiansen and Kathrine B Christensen for inputs and proof-reading.

REFERENCES

1 **Burisch J**, Jess T, Martinato M, Lakatos PL. The burden of inflammatory bowel disease in Europe. *J Crohns Colitis* 2013; **7**: 322-337 [PMID: 23395397 DOI: 10.1016/j.crohns.2013.01.010]
2 **Xavier RJ**, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* 2007; **448**: 427-434 [PMID: 17653185 DOI: 10.1038/nature06005]
3 **Khor B**, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011; **474**: 307-317 [PMID: 21677747 DOI: 10.1038/nature10209]
4 **Iskandar HN**, Ciorba MA. Biomarkers in inflammatory bowel disease: current practices and recent advances. *Transl Res* 2012; **159**: 313-325 [PMID: 22424434 DOI: 10.1016/j.trsl.2012.01.001]
5 **Ehrentraut SF**, Colgan SP. Implications of protein post-translational modifications in IBD. *Inflamm Bowel Dis* 2012; **18**: 1378-1388 [PMID: 22223542 DOI: 10.1002/ibd.22859]
6 **Poulsen NA**, Andersen V, Møller JC, Møller HS, Jessen F, Purup S, Larsen LB. Comparative analysis of inflamed and non-inflamed colon biopsies reveals strong proteomic inflammation profile in patients with ulcerative colitis. *BMC Gastroenterol* 2012; **12**: 76 [PMID: 22726388 DOI: 10.1186/1471-230X-12-76]
7 **Thompson AI**, Lees CW. Genetics of ulcerative colitis. *Inflamm Bowel Dis* 2011; **17**: 831-848 [PMID: 21319274 DOI: 10.1002/ibd.21375]
8 **Lewis JD**. The utility of biomarkers in the diagnosis and therapy of inflammatory bowel disease. *Gastroenterology* 2011; **140**: 1817-1826.e2 [PMID: 21530748 DOI: 10.1053/j.gastro.2010.11.058]
9 **Vatn MH**. Natural history and complications of IBD. *Curr Gastroenterol Rep* 2009; **11**: 481-487 [PMID: 19903424 DOI: 10.1007/s11894-009-0073-8]
10 **Meuwis MA**, Fillet M, Geurts P, de Seny D, Lutteri L, Chapelle JP, Bours V, Wehenkel L, Belaiche J, Malaise M, Louis E, Merville MP. Biomarker discovery for inflammatory bowel disease, using proteomic serum profiling. *Biochem Pharmacol* 2007; **73**: 1422-1433 [PMID: 17258689 DOI: 10.1016/j.bcp.2006.12.019]
11 **Allez M**, Lemann M, Bonnet J, Cattan P, Jian R, Modigliani R. Long term outcome of patients with active Crohn's disease exhibiting extensive and deep ulcerations at colonoscopy. *Am J Gastroenterol* 2002; **97**: 947-953 [PMID: 12003431 DOI: 10.1111/j.1572-0241.2002.05614.x]
12 **Kevans D**, Van Assche G. Is there a therapeutic window of opportunity in early inflammatory bowel disease? Early stage inflammatory bowel disease: the actual management. *Minerva Gastroenterol Dietol* 2013; **59**: 299-312 [PMID: 23867949]
13 **Bruining DH**, Siddiki HA, Fletcher JG, Tremaine WJ, Sandborn WJ, Loftus EV. Prevalence of penetrating disease and extraintestinal manifestations of Crohn's disease detected with CT enterography. *Inflamm Bowel Dis* 2008; **14**: 1701-1706 [PMID: 18623171 DOI: 10.1002/ibd.20529]
14 **Han NY**, Choi W, Park JM, Kim EH, Lee H, Hahm KB. Label-free quantification for discovering novel biomarkers in the diagnosis and assessment of disease activity in inflammatory bowel disease. *J Dig Dis* 2013; **14**: 166-174 [PMID: 23320753 DOI: 10.1111/1751-2980.12035]
15 **Yau Y**, Leong RW, Zeng M, Wasinger VC. Proteomics and metabolomics in inflammatory bowel disease. *J Gastroenterol Hepatol* 2013; **28**: 1076-1086 [PMID: 23489082 DOI: 10.1111/jgh.12193]
16 **Berry D**, Reinisch W. Intestinal microbiota: a source of novel biomarkers in inflammatory bowel diseases? *Best Pract Res Clin Gastroenterol* 2013; **27**: 47-58 [PMID: 23768552 DOI:

10.1016/j.bpg.2013.03.005]
17 **Mendoza JL**, Abreu MT. Biological markers in inflammatory bowel disease: practical consideration for clinicians. *Gastroenterol Clin Biol* 2009; **33** Suppl 3: S158-S173 [PMID: 20117339 DOI: 10.1016/S0399-8320(09)73151-3]
18 **Han NY**, Kim EH, Choi J, Lee H, Hahm KB. Quantitative proteomic approaches in biomarker discovery of inflammatory bowel disease. *J Dig Dis* 2012; **13**: 497-503 [PMID: 22988922 DOI: 10.1111/j.1751-2980.2012.00625.x]
19 **Biomarkers Definitions Working Group**. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; **69**: 89-95 [PMID: 11240971 DOI: 10.1067/mcp.2001.113989]
20 **Feng JT**, Shang S, Beretta L. Proteomics for the early detection and treatment of hepatocellular carcinoma. *Oncogene* 2006; **25**: 3810-3817 [PMID: 16799622 DOI: 10.1038/sj.onc.1209551]
21 **Marko-Varga G**, Omenn GS, Paik YK, Hancock WS. A first step toward completion of a genome-wide characterization of the human proteome. *J Proteome Res* 2013; **12**: 1-5 [PMID: 23256439 DOI: 10.1021/pr301183a]
22 **Satsangi J**, Welsh KI, Bunce M, Julier C, Farrant JM, Bell JI, Jewell DP. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. *Lancet* 1996; **347**: 1212-1217 [PMID: 8622450 DOI: 10.1016/S0140-6736(96)90734-5]
23 **Jensen ON**. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* 2004; **8**: 33-41 [PMID: 15036154 DOI: 10.1016/j.cbpa.2003.12.009]
24 **Wu L**, Candille SI, Choi Y, Xie D, Jiang L, Li-Pook-Than J, Tang H, Snyder M. Variation and genetic control of protein abundance in humans. *Nature* 2013; **499**: 79-82 [PMID: 23676674 DOI: 10.1038/nature12223]
25 **Mann M**, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003; **21**: 255-261 [PMID: 12610572 DOI: 10.1038/nbt0303-255]
26 **De Ceuleneer M**, Van Steendam K, Dhaenens M, Deforce D. In vivo relevance of citrullinated proteins and the challenges in their detection. *Proteomics* 2012; **12**: 752-760 [PMID: 22318877 DOI: 10.1002/pmic.201100478]
27 **Thelen JJ**, Miernyk JA. The proteomic future: where mass spectrometry should be taking us. *Biochem J* 2012; **444**: 169-181 [PMID: 22574775 DOI: 10.1042/BJ20110363]
28 **Hugot JP**, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411**: 599-603 [PMID: 11385576 DOI: 10.1038/35079107]
29 **Satsangi J**, Parkes M, Louis E, Hashimoto L, Kato N, Welsh K, Terwilliger JD, Lathrop GM, Bell JI, Jewell DP. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 1996; **14**: 199-202 [PMID: 8841195 DOI: 10.1038/ng1096-199]
30 **Weizman AV**, Silverberg MS. Have genomic discoveries in inflammatory bowel disease translated into clinical progress? *Curr Gastroenterol Rep* 2012; **14**: 139-145 [PMID: 22302508 DOI: 10.1007/s11894-012-0248-6]
31 **Vaiopoulou A**, Gazouli M, Theodoropoulos G, Zografos G. Current advantages in the application of proteomics in inflammatory bowel disease. *Dig Dis Sci* 2012; **57**: 2755-2764 [PMID: 22740064 DOI: 10.1007/s10620-012-2291-4]
32 **Joyner MJ**, Pedersen BK. Ten questions about systems biology. *J Physiol* 2011; **589**: 1017-1030 [PMID: 21224238 DOI: 10.1113/jphysiol.2010.201509]
33 **Nagaraj N**, Kulak NA, Cox J, Neuhauser N, Mayr K, Hoerning O, Vorm O, Mann M. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol Cell Proteomics* 2012; **11**: M111.013722 [PMID: 22021278 DOI: 10.1074/mcp.M111.013722]
34 **Steen H**, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 2004; **5**: 699-711 [PMID: 15340378 DOI: 10.1038/nrm1468]
35 **Nagaraj N**, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Pääbo S, Mann M. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* 2011; **7**: 548 [PMID: 22068331 DOI: 10.1038/msb.2011.81]
36 **Michalski A**, Cox J, Mann M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* 2011; **10**: 1785-1793 [PMID: 21309581 DOI: 10.1021/pr101060v]
37 **Cambridge SB**, Gnad F, Nguyen C, Bermejo JL, Krüger M, Mann M. Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J Proteome Res* 2011; **10**: 5275-5284 [PMID: 22050367 DOI: 10.1021/pr101183k]
38 **Thakur SS**, Geiger T, Chatterjee B, Bandilla P, Fröhlich F, Cox J, Mann M. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol Cell Proteomics* 2011; **10**: M110.003699 [PMID: 21586754 DOI: 10.1074/mcp.M110.003699]
39 **Lopes LV**, Kussmann M. Proteomics at the interface of psychology, gut physiology and dysfunction: an underexploited approach that deserves expansion. *Expert Rev Proteomics* 2011; **8**: 605-614 [PMID: 21999831 DOI: 10.1586/epr.11.50]
40 **Vermeire S**, Van Assche G, Rutgeerts P. Laboratory markers in IBD: useful, magic, or unnecessary toys? *Gut* 2006; **55**: 426-431 [PMID: 16474109 DOI: 10.1136/gut.2005.069476]
41 **Peyrin-Biroulet L**, Standaert-Vitse A, Branche J, Chamaillard M. IBD serological panels: facts and perspectives. *Inflamm Bowel Dis* 2007; **13**: 1561-1566 [PMID: 17636565 DOI: 10.1002/ibd.20226]
42 **Sandborn WJ**, Loftus EV, Colombel JF, Fleming KA, Seibold F, Homburger HA, Sendid B, Chapman RW, Tremaine WJ, Kaul DK, Wallace J, Harmsen WS, Zinsmeister AR, Targan SR. Evaluation of serologic disease markers in a population-based cohort of patients with ulcerative colitis and Crohn's disease. *Inflamm Bowel Dis* 2001; **7**: 192-201 [PMID: 11515844 DOI: 10.1097/00054725-200108000-00003]
43 **Reumaux D**, Sendid B, Poulain D, Duthilleul P, Dewit O, Colombel JF. Serological markers in inflammatory bowel diseases. *Best Pract Res Clin Gastroenterol* 2003; **17**: 19-35 [PMID: 12617880 DOI: 10.1053/bega.2002.0347]
44 **Bernstein CN**, El-Gabalawy H, Sargent M, Landers C, Rawsthorne P, Elias B, Targan SR. Assessing inflammatory bowel disease-associated antibodies in Caucasian and First Nations cohorts. *Can J Gastroenterol* 2011; **25**: 269-273 [PMID: 21647462]
45 **Peeters M**, Joossens S, Vermeire S, Vlietinck R, Bossuyt X, Rutgeerts P. Diagnostic value of anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease. *Am J Gastroenterol* 2001; **96**: 730-734 [PMID: 11280542 DOI: 10.1111/j.1572-0241.2001.03613.x]
46 **Vermeire S**, Van Assche G, Rutgeerts P. C-reactive protein as a marker for inflammatory bowel disease. *Inflamm Bowel Dis* 2004; **10**: 661-665 [PMID: 15472532 DOI: 10.1097/00054725-200409000-00026]
47 **Meuwis MA**, Vernier-Massouille G, Grimaud JC, Bouhnik Y, Laharie D, Piver E, Seidel L, Colombel JF, Louis E. Serum calprotectin as a biomarker for Crohn's disease. *J Crohns Colitis* 2013; **7**: e678-e683 [PMID: 23845231 DOI: 10.1016/j.crohns.2013.06.008]
48 **Saverymuttu SH**, Hodgson HJ, Chadwick VS, Pepys MB.

Differing acute phase responses in Crohn's disease and ulcerative colitis. *Gut* 1986; **27**: 809-813 [PMID: 3732890 DOI: 10.1136/gut.27.7.809]
49 **Solem CA**, Loftus EV, Tremaine WJ, Harmsen WS, Zinsmeister AR, Sandborn WJ. Correlation of C-reactive protein with clinical, endoscopic, histologic, and radiographic activity in inflammatory bowel disease. *Inflamm Bowel Dis* 2005; **11**: 707-712 [PMID: 16043984 DOI: 10.1097/01.MIB.0000173271.18319.53]
50 **Burri E**, Beglinger C. Faecal calprotectin in the diagnosis of inflammatory bowel disease. *Biochem Med* (Zagreb) 2011; **21**: 245-253 [PMID: 22420238 DOI: 10.11613/BM.2011.034]
51 **Tamboli CP**, Doman DB, Patel A. Current and future role of biomarkers in Crohn's disease risk assessment and treatment. *Clin Exp Gastroenterol* 2011; **4**: 127-140 [PMID: 21753895 DOI: 10.2147/CEG.S18187]
52 **Tibble J**, Teahon K, Thjodleifsson B, Roseth A, Sigthorsson G, Bridger S, Foster R, Sherwood R, Fagerhol M, Bjarnason I. A simple method for assessing intestinal inflammation in Crohn's disease. *Gut* 2000; **47**: 506-513 [PMID: 10986210 DOI: 10.1136/gut.47.4.506]
53 **Turnbaugh PJ**, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature* 2007; **449**: 804-810 [PMID: 17943116 DOI: 10.1038/nature06244]
54 **Rosenstiel P**. Stories of love and hate: innate immunity and host-microbe crosstalk in the intestine. *Curr Opin Gastroenterol* 2013; **29**: 125-132 [PMID: 23337934 DOI: 10.1097/MOG.0b013e32835da2c7]
55 **Røseth AG**, Fagerhol MK, Aadland E, Schjønsby H. Assessment of the neutrophil dominating protein calprotectin in feces. A methodologic study. *Scand J Gastroenterol* 1992; **27**: 793-798 [PMID: 1411288]
56 **Tibble JA**, Sigthorsson G, Foster R, Scott D, Fagerhol MK, Roseth A, Bjarnason I. High prevalence of NSAID enteropathy as shown by a simple faecal test. *Gut* 1999; **45**: 362-366 [PMID: 10446103]
57 **Manz M**, Burri E, Rothen C, Tchanguizi N, Niederberger C, Rossi L, Beglinger C, Lehmann FS. Value of fecal calprotectin in the evaluation of patients with abdominal discomfort: an observational study. *BMC Gastroenterol* 2012; **12**: 5 [PMID: 22233279 DOI: 10.1186/1471-230X-12-5]
58 **Kane SV**, Sandborn WJ, Rufo PA, Zholudev A, Boone J, Lyerly D, Camilleri M, Hanauer SB. Fecal lactoferrin is a sensitive and specific marker in identifying intestinal inflammation. *Am J Gastroenterol* 2003; **98**: 1309-1314 [PMID: 12818275 DOI: 10.1016/S0002-9270(03)00232-6]
59 **Angriman I**, Scarpa M, D'Incà R, Basso D, Ruffolo C, Polese L, Sturniolo GC, D'Amico DF, Plebani M. Enzymes in feces: useful markers of chronic inflammatory bowel disease. *Clin Chim Acta* 2007; **381**: 63-68 [PMID: 17368600 DOI: 10.1016/j.cca.2007.02.025]
60 **Schoepfer AM**, Trummler M, Seeholzer P, Seibold-Schmid B, Seibold F. Discriminating IBD from IBS: comparison of the test performance of fecal markers, blood leukocytes, CRP, and IBD antibodies. *Inflamm Bowel Dis* 2008; **14**: 32-39 [PMID: 17924558 DOI: 10.1002/ibd.20275]
61 **Langhorst J**, Elsenbruch S, Koelzer J, Rueffer A, Michalsen A, Dobos GJ. Noninvasive markers in the assessment of intestinal inflammation in inflammatory bowel diseases: performance of fecal lactoferrin, calprotectin, and PMN-elastase, CRP, and clinical indices. *Am J Gastroenterol* 2008; **103**: 162-169 [PMID: 17916108 DOI: 10.1111/j.1572-0241.2007.01556.x]
62 **Schoepfer AM**, Trummler M, Seeholzer P, Criblez DH, Seibold F. Accuracy of four fecal assays in the diagnosis of colitis. *Dis Colon Rectum* 2007; **50**: 1697-1706 [PMID: 17762964 DOI: 10.1007/s10350-007-0303-9]
63 **Joishy M**, Davies I, Ahmed M, Wassel J, Davies K, Sayers A, Jenkins H. Fecal calprotectin and lactoferrin as noninvasive markers of pediatric inflammatory bowel disease. *J Pediatr Gastroenterol Nutr* 2009; **48**: 48-54 [PMID: 19172123 DOI: 10.1097/MPG.0b013e31816533d3]
64 **Bunn SK**, Bisset WM, Main MJ, Gray ES, Olson S, Golden BE. Fecal calprotectin: validation as a noninvasive measure of bowel inflammation in childhood inflammatory bowel disease. *J Pediatr Gastroenterol Nutr* 2001; **33**: 14-22 [PMID: 11479402]
65 **Plevy S**, Silverberg MS, Lockton S, Stockfisch T, Croner L, Stachelski J, Brown M, Triggs C, Chuang E, Princen F, Singh S. Combined serological, genetic, and inflammatory markers differentiate non-IBD, Crohn's disease, and ulcerative colitis patients. *Inflamm Bowel Dis* 2013; **19**: 1139-1148 [PMID: 23518807 DOI: 10.1097/MIB.0b013e318280b19e]
66 **Barceló-Batllori S**, André M, Servis C, Lévy N, Takikawa O, Michetti P, Reymond M, Felley-Bosco E. Proteomic analysis of cytokine induced proteins in human intestinal epithelial cells: implications for inflammatory bowel diseases. *Proteomics* 2002; **2**: 551-560 [PMID: 11987129 DOI: 10.1002/1615-9861(200205)2:5<551::AID-PROT551>3.0.CO;2-O]
67 **Hardwidge PR**, Rodriguez-Escudero I, Goode D, Donohoe S, Eng J, Goodlett DR, Aebersold R, Finlay BB. Proteomic analysis of the intestinal epithelial cell response to enteropathogenic Escherichia coli. *J Biol Chem* 2004; **279**: 20127-20136 [PMID: 14988394 DOI: 10.1074/jbc.M401228200]
68 **Hsieh SY**, Shih TC, Yeh CY, Lin CJ, Chou YY, Lee YS. Comparative proteomic studies on the pathogenesis of human ulcerative colitis. *Proteomics* 2006; **6**: 5322-5331 [PMID: 16947118 DOI: 10.1002/pmic.200500541]
69 **Shkoda A**, Werner T, Daniel H, Gunckel M, Rogler G, Haller D. Differential protein expression profile in the intestinal epithelium from patients with inflammatory bowel disease. *J Proteome Res* 2007; **6**: 1114-1125 [PMID: 17330946 DOI: 10.1021/pr060433m]
70 **Meuwis MA**, Fillet M, Lutteri L, Marée R, Geurts P, de Seny D, Malaise M, Chapelle JP, Wehenkel L, Belaiche J, Merville MP, Louis E. Proteomics for prediction and characterization of response to infliximab in Crohn's disease: a pilot study. *Clin Biochem* 2008; **41**: 960-967 [PMID: 18489908 DOI: 10.1016/j.clinbiochem.2008.04.021]
71 **Nanni P**, Parisi D, Roda G, Casale M, Belluzzi A, Roda E, Mayer L, Roda A. Serum protein profiling in patients with inflammatory bowel diseases using selective solid-phase bulk extraction, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and chemometric data analysis. *Rapid Commun Mass Spectrom* 2007; **21**: 4142-4148 [PMID: 18022963 DOI: 10.1002/rcm.3323]
72 **Nanni P**, Mezzanotte L, Roda G, Caponi A, Levander F, James P, Roda A. Differential proteomic analysis of HT29 Cl.16E and intestinal epithelial cells by LC ESI/QTOF mass spectrometry. *J Proteomics* 2009; **72**: 865-873 [PMID: 19168159 DOI: 10.1016/j.jprot.2008.12.010]
73 **Hatsugai M**, Kurokawa MS, Kouro T, Nagai K, Arito M, Masuko K, Suematsu N, Okamoto K, Itoh F, Kato T. Protein profiles of peripheral blood mononuclear cells are useful for differential diagnosis of ulcerative colitis and Crohn's disease. *J Gastroenterol* 2010; **45**: 488-500 [PMID: 20049485 DOI: 10.1007/s00535-009-0183-y]
74 **M'Koma AE**, Seeley EH, Washington MK, Schwartz DA, Muldoon RL, Herline AJ, Wise PE, Caprioli RM. Proteomic profiling of mucosal and submucosal colonic tissues yields protein signatures that differentiate the inflammatory colitides. *Inflamm Bowel Dis* 2011; **17**: 875-883 [PMID: 20806340 DOI: 10.1002/ibd.21442]
75 **Presley LL**, Ye J, Li X, Leblanc J, Zhang Z, Ruegger PM, Allard J, McGovern D, Ippoliti A, Roth B, Cui X, Jeske DR, Elashoff D, Goodglick L, Braun J, Borneman J. Host-microbe relationships in inflammatory bowel disease detected

by bacterial and metaproteomic analysis of the mucosal-luminal interface. *Inflamm Bowel Dis* 2012; **18**: 409-417 [PMID: 21698720 DOI: 10.1002/ibd.21793]
76 **Seeley EH**, Washington MK, Caprioli RM, M'Koma AE. Proteomic patterns of colonic mucosal tissues delineate Crohn's colitis and ulcerative colitis. *Proteomics Clin Appl* 2013; **7**: 541-549 [PMID: 23382084 DOI: 10.1002/prca.201200107]
77 **Gazouli M**, Anagnostopoulos AK, Papadopoulou A, Vaiopoulou A, Papamichael K, Mantzaris G, Theodoropoulos GE, Anagnou NP, Tsangaris GT. Serum protein profile of Crohn's disease treated with infliximab. *J Crohns Colitis* 2013; **7**: e461-e470 [PMID: 23562004 DOI: 10.1016/j.crohns.2013.02.021]
78 **Alex P**, Gucek M, Li X. Applications of proteomics in the study of inflammatory bowel diseases: Current status and future directions with available technologies. *Inflamm Bowel Dis* 2009; **15**: 616-629 [PMID: 18844215 DOI: 10.1002/ibd.20652]
79 **Hall DA**, Ptacek J, Snyder M. Protein microarray technology. *Mech Ageing Dev* 2007; **128**: 161-167 [PMID: 17126887 DOI: 10.1016/j.mad.2006.11.021]
80 **Hermansson M**, Artemenko K, Ossipova E, Eriksson H, Lengqvist J, Makrygiannakis D, Catrina AI, Nicholas AP, Klareskog L, Savitski M, Zubarev RA, Jakobsson PJ. MS analysis of rheumatoid arthritic synovial tissue identifies specific citrullination sites on fibrinogen. *Proteomics Clin Appl* 2010; **4**: 511-518 [PMID: 21137068 DOI: 10.1002/prca.200900088]
81 **Tabushi Y**, Nakanishi T, Takeuchi T, Nakajima M, Ueda K, Kotani T, Makino S, Shimizu A, Hanafusa T, Takubo T. Detection of citrullinated proteins in synovial fluids derived from patients with rheumatoid arthritis by proteomics-based analysis. *Ann Clin Biochem* 2008; **45**: 413-417 [PMID: 18583628 DOI: 10.1258/acb.2007.007205]
82 **Kinloch A**, Lundberg K, Wait R, Wegner N, Lim NH, Zendman AJ, Saxne T, Malmström V, Venables PJ. Synovial fluid is a site of citrullination of autoantigens in inflammatory arthritis. *Arthritis Rheum* 2008; **58**: 2287-2295 [PMID: 18668562 DOI: 10.1002/art.23618]
83 **van Venrooij WJ**, van Beers JJ, Pruijn GJ. Anti-CCP antibodies: the past, the present and the future. *Nat Rev Rheumatol* 2011; **7**: 391-398 [PMID: 21647203 DOI: 10.1038/nrrheum.2011.76]
84 **Slade DJ**, Subramanian V, Fuhrmann J, Thompson PR. Chemical and biological methods to detect posttranslational modifications of arginine. *Biopolymers* 2013; Epub ahead of print [PMID: 23576281 DOI: 10.1002/bip.22256]
85 **Schrattenholz A**, Groebe K, Soskic V. Systems biology approaches and tools for analysis of interactomes and multi-target drugs. *Methods Mol Biol* 2010; **662**: 29-58 [PMID: 20824465 DOI: 10.1007/978-1-60761-800-3_2]
86 **Seo J**, Jeong J, Kim YM, Hwang N, Paek E, Lee KJ. Strategy for comprehensive identification of post-translational modifications in cellular proteins, including low abundant modifications: application to glyceraldehyde-3-phosphate dehydrogenase. *J Proteome Res* 2008; **7**: 587-602 [PMID: 18183946 DOI: 10.1021/pr700657y]
87 **Arnesen T**, Van Damme P, Polevoda B, Helsens K, Evjenth R, Colaert N, Varhaug JE, Vandekerckhove J, Lillehaug JR, Sherman F, Gevaert K. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc Natl Acad Sci USA* 2009; **106**: 8157-8162 [PMID: 19420222 DOI: 10.1073/pnas.0901931106]
88 **György B**, Tóth E, Tarcsa E, Falus A, Buzás EI. Citrullination: a posttranslational modification in health and disease. *Int J Biochem Cell Biol* 2006; **38**: 1662-1677 [PMID: 16730216 DOI: 10.1016/j.biocel.2006.03.008]
89 **Curis E**, Nicolis I, Moinard C, Osowska S, Zerrouk N, Bénazeth S, Cynober L. Almost all about citrulline in mammals. *Amino Acids* 2005; **29**: 177-205 [PMID: 16082501 DOI: 10.1007/s00726-005-0235-4]
90 **Ishigami A**, Ohsawa T, Hiratsuka M, Taguchi H, Kobayashi S, Saito Y, Murayama S, Asaga H, Toda T, Kimura N, Maruyama N. Abnormal accumulation of citrullinated proteins catalyzed by peptidylarginine deiminase in hippocampal extracts from patients with Alzheimer's disease. *J Neurosci Res* 2005; **80**: 120-128 [PMID: 15704193 DOI: 10.1002/jnr.20431]
91 **Wegner N**, Lundberg K, Kinloch A, Fisher B, Malmström V, Feldmann M, Venables PJ. Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. *Immunol Rev* 2010; **233**: 34-54 [PMID: 20192991 DOI: 10.1111/j.0105-2896.2009.00850.x]
92 **Klareskog L**, Stolt P, Lundberg K, Källberg H, Bengtsson C, Grunewald J, Rönnelid J, Harris HE, Ulfgren AK, Rantapää-Dahlqvist S, Eklund A, Padyukov L, Alfredsson L. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum* 2006; **54**: 38-46 [PMID: 16385494 DOI: 10.1002/art.21575]
93 **Mahdi H**, Fisher BA, Källberg H, Plant D, Malmström V, Rönnelid J, Charles P, Ding B, Alfredsson L, Padyukov L, Symmons DP, Venables PJ, Klareskog L, Lundberg K. Specific interaction between genotype, smoking and autoimmunity to citrullinated alpha-enolase in the etiology of rheumatoid arthritis. *Nat Genet* 2009; **41**: 1319-1324 [PMID: 19898480 DOI: 10.1038/ng.480]
94 **Bang SY**, Lee KH, Cho SK, Lee HS, Lee KW, Bae SC. Smoking increases rheumatoid arthritis susceptibility in individuals carrying the HLA-DRB1 shared epitope, regardless of rheumatoid factor or anti-cyclic citrullinated peptide antibody status. *Arthritis Rheum* 2010; **62**: 369-377 [PMID: 20112396 DOI: 10.1002/art.27272]
95 **Silman AJ**, Newman J, MacGregor AJ. Cigarette smoking increases the risk of rheumatoid arthritis. Results from a nationwide study of disease-discordant twins. *Arthritis Rheum* 1996; **39**: 732-735 [PMID: 8639169]
96 **Ernst A**, Jacobsen B, Østergaard M, Okkels H, Andersen V, Dagiliene E, Pedersen IS, Thorsgaard N, Drewes AM, Krarup HB. Mutations in CARD15 and smoking confer susceptibility to Crohn's disease in the Danish population. *Scand J Gastroenterol* 2007; **42**: 1445-1451 [PMID: 17852840 DOI: 10.1080/00365520701427102]
97 **Mahid SS**, Minor KS, Soto RE, Hornung CA, Galandiuk S. Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clin Proc* 2006; **81**: 1462-1471 [PMID: 17120402 DOI: 10.4065/81.11.1462]
98 **Higuchi LM**, Khalili H, Chan AT, Richter JM, Bousvaros A, Fuchs CS. A prospective study of cigarette smoking and the risk of inflammatory bowel disease in women. *Am J Gastroenterol* 2012; **107**: 1399-1406 [PMID: 22777340 DOI: 10.1038/ajg.2012.196]
99 **Lindberg E**, Tysk C, Andersson K, Järnerot G. Smoking and inflammatory bowel disease. A case control study. *Gut* 1988; **29**: 352-357 [PMID: 3356367 DOI: 10.1136/gut.29.3.352]
100 **García Rodríguez LA**, González-Pérez A, Johansson S, Wallander MA. Risk factors for inflammatory bowel disease in the general population. *Aliment Pharmacol Ther* 2005; **22**: 309-315 [PMID: 16097997 DOI: 10.1111/j.1365-2036.2005.02564.x]
101 **Tutturen AE**, Holm A, Jørgensen M, Stadtmüller P, Rise F, Fleckenstein B. A technique for the specific enrichment of citrulline-containing peptides. *Anal Biochem* 2010; **403**: 43-51 [PMID: 20399192 DOI: 10.1016/j.ab.2010.04.012]
102 **De Ceuleneer M**, De Wit V, Van Steendam K, Van Nieuwerburgh F, Tilleman K, Deforce D. Modification of citrulline residues with 2,3-butanedione facilitates their detection by liquid chromatography/mass spectrometry. *Rapid Commun*

Mass Spectrom 2011; **25**: 1536-1542 [PMID: 21594927 DOI: 10.1002/rcm.5015]
103 **Hao G**, Wang D, Gu J, Shen Q, Gross SS, Wang Y. Neutral loss of isocyanic acid in peptide CID spectra: a novel diagnostic marker for mass spectrometric identification of protein citrullination. *J Am Soc Mass Spectrom* 2009; **20**: 723-727 [PMID: 19200748 DOI: 10.1016/j.jasms.2008.12.012]
104 **Bennike T**, Lauridsen KB, Olesen MK, Andersen V, Birkelund S, Stensballe A. Optimizing the Identification of Citrullinated Peptides by Mass Spectrometry: Utilizing the Inability of Trypsin to Cleave after Citrullinated Amino Acids. *J Proteomics Bioinform* 2013; **6**: 12 [DOI: 10.4172/jpb.1000293]

P- Reviewers: Iera E, Rabago L **S- Editor**: Gou SX **L- Editor**: A
E- Editor: Ma S

Proteomics & Bioinformatics

Bennike et al., J Proteomics Bioinform 2013, 6:12
http://dx.doi.org/10.4172/jpb.1000293

Research Article Open Access

Optimizing the Identification of Citrullinated Peptides by Mass Spectrometry: Utilizing the Inability of Trypsin to Cleave after Citrullinated Amino Acids

Tue Bennike[1]*, Kasper B. Lauridsen[1], Michael Kruse Olesen[2], Vibeke Andersen[3,4], Svend Birkelund[1] and Allan Stensballe[1]

[1]Department of Health Science and Technology, Aalborg University, Fredrik Bajers Vej 3B, 9220 Aalborg, Denmark
[2]Department of Rheumatology, and Center for Clinical Research, Vendsyssel Teaching Hospital, Bispensgade 37, 9800 Hjoerring, Denmark
[3]Institute of Regional Health Services Research, University of Southern Denmark, Winslowparken 19, 3, 5000 Odense C, Denmark
[4]Organ Center, Hospital of Southern Jutland, Kresten Philipsens Vej 15, 6200 Aabenraa, Denmark

Abstract

Citrullinated proteins have been associated with several diseases and citrullination can most likely function as a target for novel diagnostic agents and unravel disease etiologies. The correct identification of citrullinated proteins is therefore of most importance. Mass spectrometry (MS) driven proteomics can with bottom up strategies analyze protein profiles and PTMs in complex samples. However, the site-specific characterization of citrullination using MS remains problematic, especially in complex samples where no sensitive chemical modification technique exists. A tryptic missed cleavage after citrulline is therefore often used as a marker for citrullination post processing. However, C-terminal tryptic citrullinated peptides have also been reported. In this study, we therefore aimed at optimizing the identification of citrullinated peptides in complex samples.

To assess the cleavage properties of trypsin, digestion was performed on synthetic peptide sets containing either arginine or citrulline. The peptide sequences originated from disease-associated *in vivo* citrullinated proteins; some reported as being C-terminal tryptic citrullinated peptides. Furthermore, the proteolytic activity was verified using digested synovial fluid samples from a rheumatoid arthritis patient. The samples were analyzed using liquid chromatography/tandem MS with electrospray ionization.

Our *in vivo* and *in vitro* studies clearly demonstrate the inability of trypsin to cleave after citrulline residues. Based on our findings, we present a strategy for verifying citrullinated sites in complex samples post processing, in proteomics shotgun experiments. By requiring a missed cleavage for the identification of citrullinated peptides, we demonstrate that 64% of false-positively annotated citrullination sites could be removed. We furthermore demonstrated likely pitfalls of applying the strategy.

In conclusion, manual annotation of citrullinated peptide spectra remains essential to ensure correct annotation. Implementing a missed cleavage requirement significantly reduces the number of spectra needing manual verification with minimal loss. This method may help future proteomics studies identify citrullinated proteins in complex samples.

Keywords: Citrulline; Citrullination; Missed cleavage; Mass spectrometry; Trypsin; Digestion; Synovial fluid

Abbreviations: Aa: Amino Acid; Anti-CCP2: Second Generation Test for Anti-Citrullinated Protein Antibodies; Arg: Arginine; Cit: Citrulline; ESI: Electrospray Ionization; FA: Formic Acid; FDR: False Discovery Rate; LC: Liquid Chromatography; m/z: Mass-to-Charge; MALDI: Matrix-Assisted Laser Desorption/Ionization; MS: Mass Spectrometry; PAD: Peptidylarginine Deiminases; PTM: Posttranslational Modification; RA: Rheumatoid Arthritis; RT: Retention Time; SP: Synthetic Peptide; XIC: Extracted Ion Chromatogram

Introduction

The modification of proteins is a common biological process. After translation of the messenger RNA into protein, most proteins are covalently modified at least once [1]. These posttranslational modifications (PTMs) are often crucial to ensure the correct physiological function of the given protein. More than 200 distinct PTMs have been identified [2]. Furthermore, the PTMs can determine the activity state, localization, turnover, and interaction with other proteins and substrates [1,3-5]. In this study, we optimized the mass spectrometry (MS) driven identification of the PTM citrullination. Citrullinated proteins and auto-antibodies against these have been associated with several diseases including: rheumatoid arthritis (RA), Alzheimer's disease, and cancer [6-8]. MS remains the only method for identifying the exact site of citrullination, however, the correct identification of citrullinated peptides from MS data by automated search engines remains troublesome [4,9].

The posttranslational modification citrullination

Citrullination is the deimination of the amino acid (aa) arginine. In the reaction, one of the side-chain nitrogen atoms of arginine is hydrolyzed, yielding the non-standard amino acid citrulline (Figure 1) along with ammonia as a side product. The reaction was first described by Fearon [10] in 1939, and citrullination can take place in

***Corresponding author:** Tue Bennike, Department of Health Science and Technology, Aalborg University, Fredrik Bajers Vej 3B, 9220 Aalborg, Denmark, Tel: +45 2613 9003; Fax: +45 9814 1818; E-mail: tbe@hst.aau.dk

Received October 27, 2013; **Accepted** December 11, 2013; **Published** December 15, 2013

Citation: Bennike T, Lauridsen KB, Olesen MK, Andersen V, Birkelund S, et al. (2013) Optimizing the Identification of Citrullinated Peptides by Mass Spectrometry: Utilizing the Inability of Trypsin to Cleave after Citrullinated Amino Acids. J Proteomics Bioinform 6: 288-295. doi:10.4172/jpb.1000293

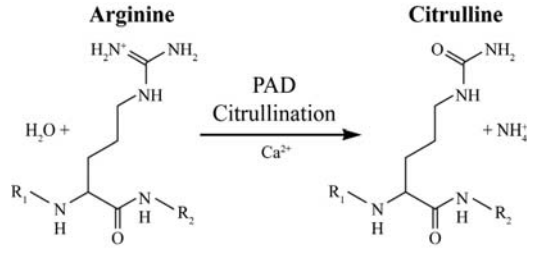


Figure 1: Citrullination of Arginine, *in vivo* catalyzed by the PAD enzyme family.

alkaline solutions at ambient temperature. Protein citrullination is *in vivo* catalyzed by a family of calcium-binding enzymes, namely the peptidylarginine deiminases (PADs). At physiological pH, arginine has a +1 charge due to the guanidine group, whereas citrulline is neutral. Thus, each citrullination event lowers the overall charge of the protein [5,11]. Essentially, the PTM leads to the loss of one positive charge and a monoisotopic mass difference of +0.984016 [4]. The loss of positive charge influences the overall charge of the protein. This in turn affects the charge distribution, isoelectric point, and hydrogen bond forming abilities of the protein [5]. The exact role of the modification remains largely unknown, but it is believed to alter the protein fold, change the protein polarity, or lead to denaturation in order to render the protein more prone to enzymatic degradation [5,11,12]. An example of a citrullinated protein is myelin basic protein, which accounts for up to 35% of the total myelin protein. Myelin is a major part of the central nervous system, where it functions as a non-conducting isolator between the nerve fibers [13,14]. Myelin exists in different charge isomers that differ by the degree of citrullination [11,15]. Finally, citrullination is known to play a role in the disease etiology of the joint disease rheumatoid arthritis (RA), and citrullinated proteins and peptides have previously been identified in the synovial fluid from these patients [12,16,17]. Moreover, RA is today diagnosed by detecting the presence of anti-citrullinated protein antibodies in the serum, with a sensitivity of 71% and a specificity of 95% [16]. The second generation test for anti-citrullinated protein antibodies (anti-CCP2) is regarded as the golden standard in diagnosing RA, however, the identification of novel disease associated citrullinated peptides may improve the anti-CCP2 test and allow for identifying clinically distinct RA patient subgroups [16]. Based on what is known about citrullinated proteins and RA, disease associated citrullinated proteins may be relevant for other inflammatory diseases as well. The knowledge of such specific citrullinated proteins will allow for analyzing disease pathways, unravel disease etiologies, and function as targets for novel diagnostic and therapeutic agents.

Detection of citrullinated peptides by mass spectrometry

In bottom up proteomic strategies, proteins are typically enzymatically digested by the protease trypsin prior to mass spectrometric analysis. Trypsin cleaves the C-terminal side of the basic aa arginine or lysine, unless either is succeeded by a proline [18]. As citrulline does not have the positive charge of arginine, the general assumption is that citrullination will result in a missed cleavage by trypsin [19-22]. This correlates well with the kinetic function of trypsin, and observations of trypsin cleavage properties [18,23]. However, tryptic peptides without a missed cleavage after citrulline have also been reported [4,24]. This could likely be due to over cleavage by trypsin or incorrect annotations made by the automated search engines. Therefore, as stated by De Ceuleneer et al. [4] in 2012, caution has to be taken when interpreting a missed cleavage as an indication of a citrullinated peptide.

In the present study, we decided to analyze the end product of tryptic digested peptides citrulline containing peptides and proteins using liquid chromatography (LC)-tandem MS sequencing. The sequences were chosen from identified disease relevant citrullinated proteins found in literature. We, furthermore, analyzed synovial fluid from a RA patient to make a similar assessment on an *in vivo* sample. This also allowed us to assess the number of citrullinated proteins found in the synovial fluid. Finally, we assessed the quality of automated citrullination annotation in the Mascot search engine, common to most proteome laboratories.

Materials and Methods

Digestion of peptides

A total of 24 synthetic peptide sets containing either arginine or citrulline (Table 1) were designed and purchased with carbamoylmethyl modified cysteine residues (JPT Peptide Technologies GmbH, Berlin, Germany). The aa sequences originated from reported citrullinated tryptic peptides added to 0.05 µg from *in vivo* modified proteins, and some sites were reported as C-terminal citrullinated. The peptide sequences were selected to ensure that nearly all peptides had at least one lysine or arginine to confirm the successful digestion. The freeze-dried peptides were resuspended according to the manufacturer's instructions, and the masses were verified by matrix assisted laser desorption/ionization (MALDI) MS (data not shown). Tryptic digestion was performed in 5% acetonitrile (ACN) and 50 mM ammonium bicarbonate. 0.5 µg crudepeptide was added to 0.05 sequencing grade trypsin (Promega, Fitchburg, USA). The samples were digested overnight at 37°C and acidified with formic acid (FA) to a concentration of 5% followed by nanoLC-MS/MS analysis. The undigested samples were treated identically, but the addition of trypsin was omitted. The peptides were LC-MS analyzed one sample at a time, except in one case where two samples were injected simultaneously to verify the observed change in retention time (RT). All mass-calculations were performed in GPMAW 9.02 (Lighthouse Data, Odense, Denmark).

Synovial fluid sample preparation

Synovial fluid was obtained from a 69 year old female RA patient, who was tested positive for anti-cyclic citrullinated peptide antibodies and rheumatoid factor. The digestion of synovial fluid prior to MS analysis was performed using the filter aided sample preparation Protein Digestion Kit (Expedeon, San Diego, USA) according to manufacturer's instructions using 30 kDa cutoff spin filters. 90 µg total SF protein was digested using two µg sequencing grade modified trypsin (Promega), and the samples were digested overnight at 37°C. After trypsin digestion, the samples were acidified with trifluoroacetic acid (TFA), desalted with TAGRA C18 columns (Nest Group, Southborough, USA), and finally resuspended in 2% acetonitrile (ACN)1% FA.

Mass spectrometry

The synthetic peptide samples were analyzed by automated LC-electrospray ionization (ESI) MS/MS using a 1200 series Agilent nanoflow HPLC (Agilent Technologies, Santa Clara, USA) system coupled online by a nanospray ion source (Proxeon, Odense, Denmark)

SP#	Protein Name	Protein ID	Sequence	Ref.	% RT shift
1	α-enolase	P06733.2	13-DS **R/Cit** GNPTVEVDLFTS**K**GLF**R**-32	[20]	2.4
2	-	P06733.2	266-DPS **R/Cit** YISPDQLADLY**K**SFI**K**-285	[20]	3.0
3	-	P06733.2	415-EELGS**KAK**FAG **R/Cit** NF **R/Cit** NPLA**K**-434	[20]	1.6
4	Vimentin	AAH66956	31-VTTST**R**TYSLGSAL **R/Cit** PSTS**R**-50	[21]	1.4
5	-	AAH66956	136-EQL**K**GQG**K**S **R/Cit** LGDLYEEEM**R**-155	[21]	2.6
6	-	AAH66956	371-NM**K**EEMA**R**HL **R/Cit** EYQDLLNV**K**-390	[21]	1.4
7	-	AAH66956	371-NM**K**EEMA **R/Cit** HL**R**EYQDLLNV**K**-390	[21]	2.0
8	protein-arginine deiminase type-4	NP_036519.2	201-LHVA **R/Cit** SEMD**K**V **R/Cit** VFQAT **R/Cit** G**K**-220	[30]	3.6
9	-	NP_036519	375-GL**K**EFPI**K** **R/Cit** VMGPDFGYVT**R**-394	[30]	3.4
10	-	NP_036519	480-PAPD**RK**GF**R**LLLASP **R/Cit** SCY**K**-499	[30]	2.0
11	C-X-C motif chemokine 10	P02778.2	24-LS **R/Cit** TV**R**CTCISISNQPVNP**R**-43	[31]	1.6
12	tubulin polymerization promoting protein p25 α	ACB10579	72-EMHG**K**NWS**K**LC **R/Cit** DCQVIDG**R**-91	[22]	1.6
13	-	ACB10579	154-SGVT**K**AISSPTVS **R/Cit** LTDTT**K**-173	[22]	1.4
14	Fibrinogen α-chain	P02671	29-AEGGGV **R/Cit** GP**R**VVE **R/Cit** HQSAC**K**-48	[32]	0
15	-	P02671	562-SHHPGIAEFPS **R/Cit** G**K**SSSYS**K**-581	[32]	0.4
16	-	P02671	583-FTSSTSYN **R/Cit** GDSTFES**K**SY**K**-602	[32]	1.0
17	Myelin basic protein	P02686.3	158-A **R/Cit** HGFLP **R/Cit** H**R**DTGILDSIG**R**-177	[33]	2.6
18	-	P02686.3	245-LS**R**FSWGAEGQ **R/Cit** PGFGYGG**R**-264	[33]	2.0
19	-	P02686.3	257-PGFGYGG **R/Cit** ASDY**K**SAH**K**GF**K**-276	[33]	0.6
20	-	P02686.3	284-LS**K**IF**K**LGG **R/Cit** DS**R**SGSPMA**R**-303	[33]	0.6
21	Fibrinogen β-chain	P02675	266-Y **R/Cit** VYCDMNTENGGWTVIQN**R**-285	[24]	0
22	-	P02675	254-MYLIQPDSSV**K**PY **R/Cit** VYCDM**R**-273	[24]	3.4
23	-	P02675	55-EAPSL **R/Cit** PAPPPISGGGY**RA**R-74	[24]	2.0
24	Fibrinogen γ-chain	P02679.3	132-SI **R/Cit** YLQEIYNSNNQ**K**IVNL**K**-151	[24]	3.0

Table 1: The complete list of synthetic peptides (SP) derived from human proteins, along with protein accession number, amino acid position, the reference to the citrullinated protein, and the observed RT shift between the citrulline and arginine version of the peptide. Bold letters indicates expected tryptic cleavage sites, and Cit indicates citrulline.

to a hybrid microQTOF mass spectrometer (Bruker, Bremen, Germany). One pmol sample was loaded on an in-house packed 10 cm reversed phase C18 column using a single column system (Dr. Maisch, Germany; Reprosil-pur C18-AQ). The sample was eluted with a linear gradient of 98% solvent A (0.1% FA, 0.005% heptafluorobutyric acid) and 2% solvent B (90% ACN, 0.1% FA, 0.005% heptafluorobutyric acid), which was increased to 40% solvent B on a 30 minutes ramp gradient at a constant flow rate of 200 nL/min. The mass spectrometer was used in data dependent mode to automatically switch between MS and MS/MS acquisition, selecting the three most abundant ions. The resulting raw files were analyzed using Bruker Daltonics Data Analysis v 3.4 (Build 192), and extracted ion chromatograms (XIC) were constructed with all predicted tryptic peptides ± m/z 0.01, under the assumption that trypsin cleaves after arginine, lysine, and citrulline.

The synovial fluid sample was analyzed on an automated LC-ESI MS/MS setup using an UltiMate™ 3000 UPLC system upgraded with a RSLC nanopump module. The system was coupled online with an emitter for nanospray ionization (New objective 360-20-10) to a Q Exactive mass spectrometer (Thermo Scientific, Waltham, USA). Five µg sample were loaded onto a C18 reversed phase column (Dionex Acclaim PepMap RSLC C18, 2 µm, 100 Å, 75 µm×50 cm) and eluted with a linear gradient of 96% solvent A (0.1% FA, 0.005% heptafluorobutyric acid) and 4% solvent B (90% ACN, 0.1% FA, 0.005% heptafluorobutyric acid), which was increased to 35% solvent B on a 120 minutes ramp gradient at a constant flow rate of 300 nL/min. The mass spectrometer was used in a data dependent mode, selecting the 12 ions with the highest intensity for HCD fragmenting. Fragmented ions were dynamically excluded for 30 sec. The resulting raw files were analyzed using Thermo Proteome Discoverer v 1.4 (build 288) connected to a Mascot server with a local Homo sapiens Uniprot database. The raw files were searched against the database using the following parameters: maximum three missed cleavages, 10 ppm precursor mass tolerance, 30 mmu fragment mass tolerance, variable modifications: Citrullination (R), Deamidation (NQ), Oxidation (M), static modifications: carboxymethyl (C), and percolator with 1% false discovery rate (FDR). All peptides annotated as citrullinated or deamidated were subsequently manually assessed. For a successful annotation of citrullination, the PTM should either be visible on the fragment ion, or by the subsequent fragment ion series. Raw files were inspected in Qual Browser Thermo Xcalibur v 2.2 (build 42).

Results and Discussion

Tryptic cleavage of citrulline containing peptides

To demonstrate the overall effectiveness of the protease trypsin to cleave at modified residues, we investigated a total of 24 peptide pairs (Table 1). In the first set of experiments, the theoretical outcome by *in situ* digestion were examined and compared to the empirical data. Protein-arginine deiminase type-4 can be citrullinated at position 495_{Arg}. A synthetic peptide pair SP 10 and SP 10_{Cit} with arginine or citrulline respectively were made with the aa sequence 480-PAPD**RK**GF**R**LLLASP**R/Cit**SCY**K**-499 (Table 1). The digestion experiment demonstrated that trypsin cleaves after arginine, but not citrulline (Figure 2). For SP 10, prior to digestion, only the synthetic peptide was detected in the XIC. After digestion, peptides corresponding to PAPD**R**, LLLASP**R** and SCY**K** were detected, corresponding to a complete cleavage after 495_{Arg} (Figure 3). This was not the case for the citrullinated peptide, SP 10_{Cit}, where peptides corresponding to PAPD**R** and LLLASP**Cit**SCY**K** were detected after digestion. All investigated peptides demonstrated this behavior, and tryptic cleavage after a citrulline residue was never observed.

Commonly, in solution trypsin digestion protocols, a trypsin-

to-protein ratio of 1:50 to 1:100 is employed and a digestion time varying from a few hours to overnight digestion [18]. To ensure a complete tryptic digestion, we employed a significantly higher trypsin concentration in the performed tryptic digestions (1:10) combined with 12 h incubation at 37°C [25]. However, cleavage after citrulline could not be detected for any of the investigated peptides, whereas the arginine and lysine cleavage sites were fully cleaved. Our study thereby demonstrates that tryptic cleavage after citrulline is not occurring to a detectable extent under standard digestion conditions.

Citrullination and retention time shift

Citrullination results in an increased hydrophobicity of the modified peptide, and as an outcome citrullinated peptides will have a longer RT on a reversed phase high-performance LC column [26]. The XIC of the coinjection of SP 3 and $SP3_{cit}$ (Figure 4) demonstrates that the RT shift between these two peptides on this system and gradient is 3.0%. The RT shifts for the peptide sets were extracted across different LC runs (Table 1). For 22 of the 24 peptide sets, the RT shift was greater than 0.4% (7 s on the 30 min gradient), which in this experiment was enough to ensure baseline separation. The results demonstrate that coelusion for most peptides will not occur, which confirms the findings by Raijmakers et al. [26]. This becomes important when considering the +0.984016 Da mass shift caused by citrullination (and deamidation) which is close to that of a single neutron +1.008665 Da. As a result, if a citrulline and an arginine containing peptide are coeluting, the m/z signal from the citrullinated peptide will fall within the isotopic cluster of the non-citrullinated peptide. This will cause an apparent change in the isotopic cluster [27]. Due to the typical isolation window used on mass spectrometers (m/z 1 to 3), both the modified and unmodified peptide would in this scenario be selected for fragmentation, which most likely would hinder a successful identification of the PTM [4]. An alternative solution to dealing with coelusion would be to run the samples on a MS with ion-mobility, as this would likely separate the citrullinated peptides from the non-citrullinated peptides regardless of LC separation.

In a tryptic digest containing citrullinated proteins and non-citrullinated proteins, a likely scenario in *in vivo* samples, the site of citrullination will in the non-modified peptide be a tryptic cleavage site. The mass of the citrullinated peptide will, therefore, in most cases not overlap with the mass of the non-citrullinated peptide. However, when using other proteases such as LysC the RT shift becomes important for the successful identification of both peptides [4]. This is relevant when assessing the ratio of citrullinated peptides, as e.g. LysC will result in two comparable peptides. De Ceuleneer et al. [27] demonstrated in 2012 that the degree of skewed isotope pattern can be used to quantify the amount of citrullinated peptide, by designing gradients where the two peptides do coelute. However, in order to have an overlap of elusion, the LC gradient must be designed accordingly as at least 22 of our 24 peptide sets did not fulfill this requirement on our setup. This might limit the feasibility of the approach for complex samples.

Tryptic missed cleavage as a citrullination verification strategy

The identical mass shift of deamidation of asparginine or glutamine causes it to be mistaken for citrullination and vice versa by automated protein identification search engines, such as Mascot (Matrix Science, Boston, USA). This is especially pronounced if only one of the modifications is allowed as a variable modification in the search (data not shown). Resultantly, a manual verification of citrullinated peptide candidate tandem mass spectra remains necessary for unambiguous site specific identification. However, our findings support that a missed cleavage can be used to significantly reduce the number of reported citrullinated peptide spectra, which needs to be manually verified, since a deamidation of asparagine or glutamine will not lead to a tryptic missed cleavage. Combining the correct mass increment with the MS/

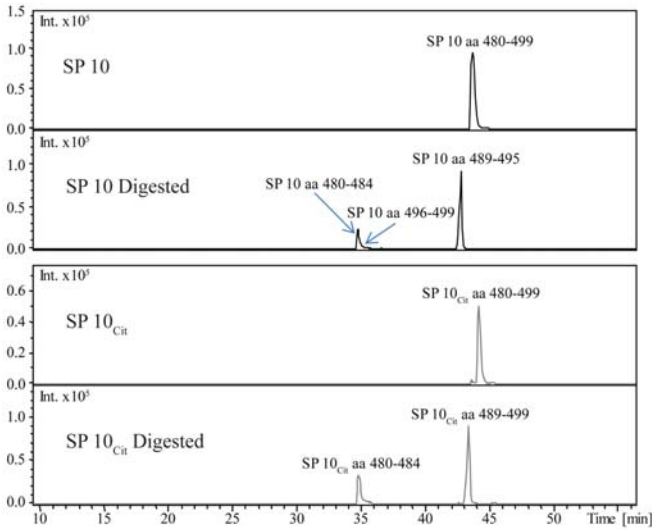


Figure 2: XIC of predicted citrullinated and non-citrullinated tryptic peptides +/- m/z 0.01, before and after tryptic digestion of SP 10 and $SP\ 10_{Cit}$, with the sequence aa 480-499. For digested SP 10, the fragments aa 489-495 and aa 496-499 are detected, confirming a cleavage after aa 495_{Arg}. For digested $SP\ 10_{cit}$, the fragment aa 489-499 is detected, confirming a missed cleavage after aa 495_{Cit}.

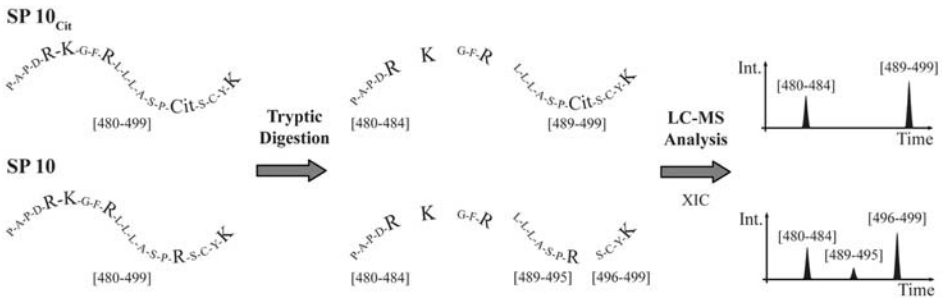


Figure 3: Citrullination protects the natural cleavage site in SP 10_{Cit}, where the entire fragment aa 489-499 is detected post digestion. For SP 10, the fragments aa 489-499 is cleaved into aa 489-495 and aa 496-499.

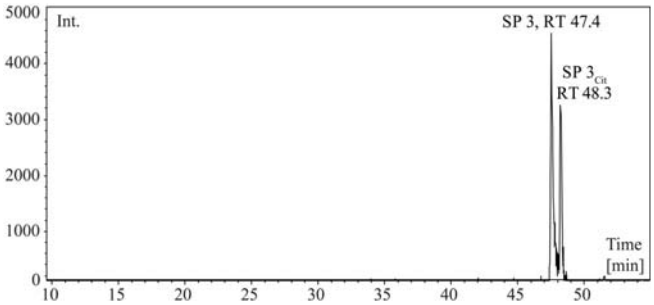


Figure 4: XIC of a co-injection of SP 3 and SP 3_{Cit} +/- m/z 0.01 demonstrating the RT shift caused by citrullination. The RT shift is calculated to be 3.0% (0.9 min) within this 30 minute gradient.

MS information, RT shift (if the non-citrullinated peptide is available) and hindered cleavage by trypsin seems to be a valid approach for a reliable identification of citrullinated peptides. It should be noted that *in vivo* samples may contain tryptic peptides with a citrullinated C-terminal, if the C-terminal of the original protein ends on a modified arginine. Using the above described identification parameters these peptides will be dismissed as false positives.

To verify the *in vitro* findings, synovial fluid from a RA patient was analyzed. The data was processed using Proteome Discoverer with Mascot as search engine, and a total of 364 proteins were identified at 1% FDR. All peptides annotated as being deamidated (NQ) or citrullinated (R) and contained at least one arginine were inspected.

Ten peptides were annotated as being deamidated (and not citrullinated), and contained at least one arginine. Of these, one was found to actually be citrullinated.

A total of 58 peptides were annotated as being citrullinated. 37 of these 58 peptides (64%) could be annotated directly and unambiguously from the fragment mass, and 7 of the 58 peptides (12%) by the masses of flanking fragments. However, 14 of the 58 peptides (24%) were found not to be citrullinated (Figure 5A).

When investigating non-citrulline peptides that were incorrectly annotated in greater detail, 10 of the 14 peptides (64%) were marked as having a citrullinated arginine on the C-terminal (Figure 5B). The remaining 4 of the 14 peptides (36%) contained an internal arginine residue. The detected mass shift was a result of the secondary isotope peak having been selected for MS/MS fragmentation, and not the monoisotopic peak. This was also the case for some of the peptides being incorrectly annotated as C-terminal citrullinated. The secondary isotope leads to a +1.008665 Da mass shift, which can be mistaken for a citrullination or a deamidation, especially in low resolution fragment spectra.

One of the 44 correctly annotated citrullinated peptides (2%) contained a citrullination on the C-terminal. Upon further analysis, the peptide was identified as the Complement C3f fragment (PMID P01024). The citrullination was found to be on position 1320, the C-terminal aa of the C3f fragment. C3f is a peptide that is released *in vivo* when C3b is converted to C3ib by the serine protease Factor I during complement activation [28,29]. Therefore, it seems likely that this is an example of the citrullinations that might be missed when implementing the no C-terminal citrullination limitation. Thus, such peptides will be lost in the data analysis when searching for citrullinated peptides and not allowing C-terminal citrullinations.

Finally, 30 of the 37 citrullinated peptides that could be unambiguously annotated, belonged to proteins previously identified as being citrullinated in synovial fluid by Van Beers et al. [17] (Table 2). Seven of the synovial fluid proteins have to the best of our knowledge not prior been identified as being citrullinated.

C-terminal peptides such as C3f are present; however, as demonstrated, by far the majority of the C-terminal citrullinated

peptides identified were incorrectly annotated. It therefore seems plausible that many peptides assigned as being C-terminal citrullinated in literature are due to deamidations being mistaken for citrullinations or that the secondary isotope peak of the peptide was selected for fragmentation. Hence, a manual verification of the fragment spectra seems necessary if a C-terminal citrulline is to be confidently annotated. Another means of unambiguously identifying a citrullination, is to use a chemical modification, as described by De Ceuleneer et al. [30].

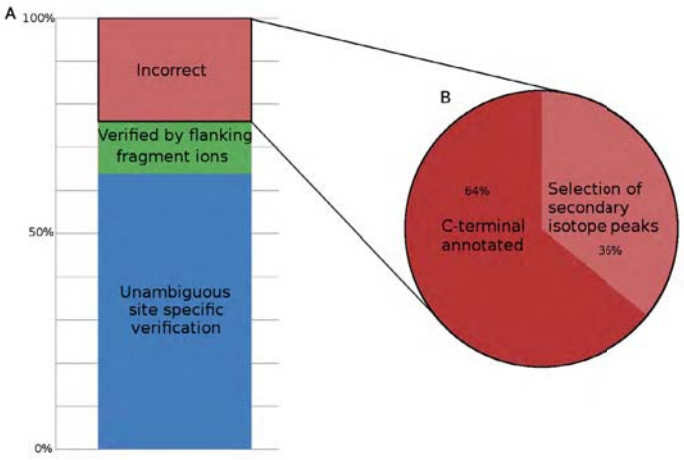


Figure 5: A: Distribution result of the manual PTM analysis of the peptides, annotated as being citrullinated by Mascot. **B:** Distribution of cause for incorrect citrullination annotation by Mascot.

Protein name	Protein ID	Sequence	Mod.Res.
Actin, cytoplasmic 1	P60709	85-IWHHTFYNEL **Cit** VAPEEHPVLLTEAPLNPK-113	96
α-1-antitrypsin	P01009	218-WE **Cit** PFEVKDTEEEDFHVDQVTTVK-241	220
α-2-macroglobulin	P01023	704-VGFYESDVMG **Cit** GHAR-719	715
Apolipoprotein A-I	P02647	231-AKPALEDL **Cit** QGLLPVLESFK-250	239
Apolipoprotein A-IV	P06727	276-GNL **Cit** GNTEGLQK-287	279
Apolipoprotein C-III	P02656	45-DALSSVQESQVAQQA **Cit** GWVTDGFSSLK-71	60
Apolipoprotein E	P02649	186-EGAE **Cit** GLSAIR-198	190
Apolipoprotein L1	O14791	321-VNEPSILEMS **Cit** GVK-334	331
Complement C1r subcomponent	P00736	472-MGNFPWQVFTNIHG **Cit** GGGALLGDR-494	484
Complement C3	P01024	1285-DAPDHQELNLDVSLQLPS **Cit**-1303	1303
Complement C4-A	P0C0L4	1350-QI **Cit** GLEEELQFSLGSK-1365	1352
Fibrinogen α-chain	P02671	582-QFTSSTSYN **Cit** GDSTFESK-599	591
Fibrinogen α-chain	P02671	481-EVVTSEDGSDCPEAMDLGTLSGIGTLDGF **Cit** HR-512	510
Fibrinogen α-chain	P02671	259-MELERPGGNEIT **Cit** GGSTSYGTGSETESPR-287	272
Fibrinogen α-chain	P02671	115-GDFSSANN **Cit** DNTYNR-129	123
Fibrinogen β-chain	P02675	53-**Cit** EEAPSLRPAPPPISGGGYR-72	53
Fibronectin	P02751	253-GNLLQCICTGNG **Cit** GEWK-269	265
Fibronectin	P02751	2335-RPGGEPSPEGTTGQSYNQYSQ **Cit** YHQR-2360	2354
Haptoglobin	P00738	117-L **Cit** TEGDGVYTLNNEK-131	118
Ig κ-chain V-III region GOL	P04206	46-LLMYGASS **Cit** ATGIPDRFSGSGSGTDFTLTISR-78	55
Inter-α-trypsin inhi-bitor heavy chain H2	P19823	136-TVG **Cit** ALYAQAR-146	139
Inter-α-trypsin inhi-bitor heavy chain H4	Q14624	658-MNF **Cit** PGVLSSR-669	661
Plasminogen	P00747	88-M **Cit** DVVLFEK-96	89
Protein AMBP	P02760	294-GPC **Cit** AFIQLWAFDAVK-309	297
Proteoglycan 4	Q92954	187-NSAAN **Cit** ELQK-196	191
Proteoglycan 4	Q92954	1307-AIGPSQTHTI **Cit** IQYSPARLAYQDK-1330	1317
Proteoglycan 4	Q92954	1285-RPALNYPVYGETTQV **Cit** R-1301	1300
Prothrombin	P00734	453-YNW **Cit** ENLD **Cit** DIALMK-467	456,461
Prothrombin	P00734	434-YE **Cit** NIEK-440	436
Serum albumin	P02768	97-LCTVATL **Cit** ETYGEMADCCAK-117	105

Table 2: The complete list of proteins where citrulline has been annotated directly and unambiguously on the fragment. Bold protein name indicates newly identified citrullinated RA synovial fluid proteins. All other proteins have been found by van Beers JJ et al. [17] to contain citrulline.

The citrulline residue is modified by the addition of an imidazolone derivative, which results in a mass shift of +50 Da. This method has the advantage of being highly specific for citrulline residues however, as noted in the article the method requires a high concentration of the citrullinated peptide(s) and the method is not suitable for complex protein samples. The missed cleavage strategy is applied post acquisition and search hence no chemical modification is required and the method is compatible with standard shotgun proteomics on complex samples.

Conclusion

Citrullination has been associated with several diseases, and auto-antibodies against citrullinated proteins are today used as an important clinical diagnostic biomarker for characterizing rheumatoid arthritis. The exact physiological role of citrullination in relation to diseases is incomplete, and specific analyses are needed to expand upon current knowledge. However, the essential MS driven verification of the exact site of modification remains problematic.

Using 24 sets of synthetic peptides containing citrulline and arginine, with sequences from previously identified citrullinated proteins reported in the literature, we have demonstrated the inability of trypsin to cleave after a citrulline residue. Furthermore, our study confirms that the RT shift between a citrullinated and a non-citrullinated peptide on a short reversed phase (C18) column and a 30 minutes gradient in most cases is enough to ensure that the two peptides will not coelute. C-terminal citrullinated tryptic peptides can, therefore, only occur if a given protein or protein fragment ends on a citrullinated arginine residue, which is the case for the citrullinated protein fragment C3f. Deamidation of asparagine or glutamine will not lead to a missed cleavage. Therefore, a missed cleavage can be used as a marker to reduce the number of reported citrullinated peptide spectra which needs to be manually validated. In an *in vivo* synovial fluid sample from a RA patient, 64% of the false positively reported citrullinations could be readily dismissed using this strategy. In addition, 7 proteins, not previously reported as being citrullinated, have been identified in the RA synovial fluid.

References

1. Mann M, Jensen ON (2003) Proteomic analysis of post-translational modifications. Nat Biotechnol 21: 255-261.
2. Thelen JJ, Miernyk JA (2012) The proteomic future: where mass spectrometry should be taking us. Biochem J 444: 169-181.
3. Jensen ON (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. Curr Opin Chem Biol 8: 33-41.
4. De Ceuleneer M, Van Steendam K, Dhaenens M, Deforce D (2012) In vivo relevance of citrullinated proteins and the challenges in their detection. Proteomics 12: 752-760.
5. György B, Tóth E, Tarcsa E, Falus A, Buzás EI (2006) Citrullination: a posttranslational modification in health and disease. Int J Biochem Cell Biol 38: 1662-1677.
6. Schellekens GA, de Jong BA, van den Hoogen FH, van de Putte LB, van Venrooij WJ (1998) Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. J Clin Invest 101: 273-281.
7. Ishigami A, Ohsawa T, Hiratsuka M, Taguchi T, Kobayashi S, et al. (2005) Abnormal accumulation of citrullinated proteins catalyzed by peptidylargininedeiminase in hippocampal extracts from patients with Alzheimer's disease. J Neurosci Res 80: 120-128.
8. Chang X, Fang K (2010) PADI4 and tumourigenesis. Cancer Cell Int 10: 7.
9. Slade DJ, Subramanian V, Fuhrmann J, Thompson PR (2013) Chemical and biological methods to detect posttranslational modifications of arginine. Biopolymers.
10. Fearon WR (1939) Thecarbamidodiacetyl reaction: a test for citrulline. Biochem J 33: 902-907.
11. Curis E, Nicolis I, Moinard C, Osowska S, Zerrouk N, et al. (2005) Almost all about citrulline in mammals. Amino Acids 29: 177-205.
12. Hermansson M, Artemenko K, Ossipova E, Eriksson H, Lengqvist J, et al. (2010) MS analysis of rheumatoid arthritic synovial tissue identifies specific citrullination sites on fibrinogen. Proteomics ClinAppl 4: 511-518.
13. Hartline DK, Colman DR (2007) Rapid conduction and the evolution of giant axons and myelinated fibers. Curr Biol 17: R29-35.
14. Sjostrand FS (1953) Thelamellated structure of the nerve myelin sheath as revealed by high resolution electron microscopy. Experientia 9: 68-69.
15. Kim JK, Mastronardi FG, Wood DD, Lubman DM, Zand R, et al. (2003) Multiple sclerosis: an important role for post-translational modifications of myelin basic protein in pathogenesis. Mol Cell Proteomics 2: 453-462.
16. vanVenrooij WJ, van Beers JJ, Pruijn GJ (2011) Anti-CCP antibodies: the past, the present and the future. Nat Rev Rheumatol 7: 391-398.
17. van Beers JJ, Schwarte CM, Stammen-Vogelzangs J, Oosterink E, Božič B, et al. (2013) The rheumatoid arthritis synovial fluid citrullinome reveals novel citrullinated epitopes in apolipoprotein E, myeloid nuclear differentiation antigen, and ß-actin. Arthritis Rheum 65: 69-80.
18. Olsen JV, Ong SE, Mann M (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics 3: 608-614.
19. Kinloch A, Tatzer V, Wait R, Peston D, Lundberg K, et al. (2005) Identification of citrullinated alpha-enolase as a candidate autoantigen in rheumatoid arthritis. Arthritis Res Ther 7: R1421-1429.
20. Bang H, Egerer K, Gauliard A, Lüthke K, Rudolph PE, et al. (2007) Mutation and citrullination modifies vimentin to a novel autoantigen for rheumatoid arthritis. Arthritis Rheum 56: 2503-2511.
21. Kleinnijenhuis AJ, Hedegaard C, Lundvig D, Sundbye S, Issinger OG, et al. (2008) Identification of multiple post-translational modifications in the porcine brain specific p25alpha. J Neurochem 106: 925-933.
22. Ordóñez A, Martínez-Martínez I, Corrales FJ, Miqueo C, Miñano A, et al. (2009) Effect of citrullination on the function and conformation of antithrombin. FEBS J 276: 6763-6772.
23. Vandermarliere E, Mueller M, Martens L (2013) Getting intimate with trypsin, the leading protease in proteomics. Mass Spectrom Rev 32: 453-465.
24. Zhao X, Okeke NL, Sharpe O, Batliwalla FM, Lee AT, et al. (2008) Circulating immune complexes contain citrullinated fibrinogen in rheumatoid arthritis. Arthritis Res Ther 10: R94.
25. Hustoft HK, Reubsaet L, Greibrokk T, Lundanes E, Malerod H (2011) Critical assessment of accelerating trypsination methods. J Pharm Biomed Anal 56: 1069-1078.
26. Raijmakers R, van Beers JJ, El-Azzouny M, Visser NF, BožičB, et al. (2012) Elevated levels of fibrinogen-derived endogenous citrullinated peptides in synovial fluid of rheumatoid arthritis patients. Arthritis Res Ther 14: R114.
27. De Ceuleneer M, Van Steendam K, Dhaenens M, Elewaut D, Deforce D (2012) Quantification of citrullination by means of skewed isotope distribution pattern. J Proteome Res 11: 5245-5251.
28. Harrison RA, Lachmann PJ (1980) The physiological breakdown of the third component of human complement. Mol Immunol 17: 9-20.
29. Ganu VS, Müller-Eberhard HJ, Hugli TE (1989) Factor C3f is a spasmogenic fragment released from C3b by factors I and H: the heptadeca-peptide C3f was synthesized and characterized. Mol Immunol 26: 939-948.
30. De Ceuleneer M, De Wit V, Van Steendam K, Van Nieuwerburgh F, Tilleman K, et al. (2011) Modification of citrulline residues with 2,3-butanedione facilitates their detection by liquid chromatography/mass spectrometry. Rapid Commun Mass Spectrom 25: 1536-1542.
31. Andrade F, Darrah E, Gucek M, Cole RN, Rosen A, et al. (2010) Autocitrullination of human peptidyl arginine deiminase type 4 regulates protein citrullination during cell activation. Arthritis Rheum 62: 1630-1640.
32. Loos T, Mortier A, Gouwy M, Ronsse I, Put W, et al. (2008) Citrullination of CXCL10 and CXCL11 by peptidylargininedeiminase: a naturally occurring posttranslational modification of chemokines and new dimension of immunoregulation. Blood 112: 2648-2656.

33. Kubota K, Yoneyama-Takazawa T, Ichikawa K (2005) Determination of sites citrullinated by peptidylargininedeiminase using 18O stable isotope labeling and mass spectrometry. Rapid Commun Mass Spectrom 19: 683-688.

34. Wood DD, Moscarello MA (1989) The isolation, characterization, and lipid-aggregating properties of a citrulline containing myelin basic protein. J Biol Chem 264: 5121-5127.

Citation: Bennike T, Lauridsen KB, Olesen MK, Andersen V, Birkelund S, et al. (2013) Optimizing the Identification of Citrullinated Peptides by Mass Spectrometry: Utilizing the Inability of Trypsin to Cleave after Citrullinated Amino Acids. J Proteomics Bioinform 6: 288-295. doi:10.4172/jpb.1000293

Journal of proteome research

Article

pubs.acs.org/jpr

Open Access on 08/26/2015

A Normative Study of the Synovial Fluid Proteome from Healthy Porcine Knee Joints

Tue Bennike,[†,⊥] Ugur Ayturk,[‡,#] Carla M. Haslauer,[‡] John W. Froehlich,[§] Benedikt L. Proffen,[‡] Omar Barnaby,[†] Svend Birkelund,[⊥] Martha M. Murray,[‡] Matthew L. Warman,[‡,#,∥] Allan Stensballe,[⊥] and Hanno Steen*[,†]

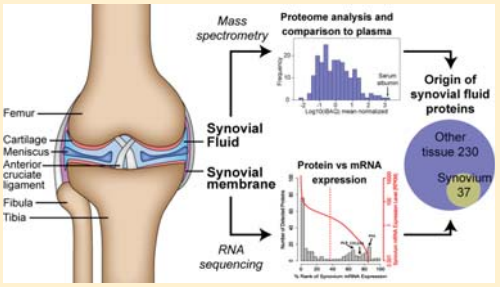
[†]Department of Pathology and Proteomics Center, [‡]Department of Orthopaedic Surgery, [§]Department of Urology, and [∥]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, Massachusetts 02115, United States

[⊥]Department of Health Science and Technology, Aalborg University, Aalborg DK-9220, Denmark

[#]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, United States

Supporting Information

ABSTRACT: Synovial fluid in an articulating joint contains proteins derived from the blood plasma and proteins that are produced by cells within the joint tissues, such as synovium, cartilage, ligament, and meniscus. The proteome composition of healthy synovial fluid and the cellular origins of many synovial fluid components are not fully understood. Here, we present a normative proteomics study using porcine synovial fluid. Using our optimized method, we identified 267 proteins with high confidence in healthy synovial fluid. We also evaluated mRNA expression data from tissues that can contribute to the synovial fluid proteome, including synovium, cartilage, blood, and liver, to better estimate the relative contributions from these sources to specific synovial fluid components. We identified 113 proteins in healthy synovial fluid that appear to be primarily derived from plasma transudates, 37 proteins primarily derived from synovium, and 11 proteins primarily derived from cartilage. Finally, we compared the identified synovial fluid proteome to the proteome of human plasma, and we found that the two body fluids share many similarities, underlining the detected plasma derived nature of many synovial fluid components. Knowing the synovial fluid proteome of a healthy joint will help to identify mechanisms that cause joint disease and pathways involved in disease progression.



KEYWORDS: *Synovial fluid, synovium, plasma, porcine, human, proteomics, transcriptomics, origin, PTM, method optimization*

INTRODUCTION

Synovial fluid is present in all joint cavities, where it protects the articular cartilage surfaces, in part by reducing friction. Synovial fluid, furthermore, facilitates the transport of nutrients and waste products including proteins and metabolites between the vascularized synovium and the avascular cartilage.[1−3] Many components of synovial fluid are derived from blood plasma, and these two body fluids share many similarities in terms of protein composition.[4,5] However, synovial fluid also contains proteins secreted from the surrounding tissue, including the articular cartilage and synovium.[6] The protein concentration in synovial fluid from healthy knee joints is approximately 25 mg/mL, i.e., ∼$^1/_3$ of the concentration found in blood plasma, and albumin constitutes approximately 12 mg/mL.[1,3,7−10]

Joint diseases, in particular osteoarthritis (OA) and rheumatoid arthritis (RA), are the leading cause of disability in people over 55 years.[6] It has been estimated that, as of 2005, 27 million adults in the United States have clinical OA, and in 2009, OA was the fourth most common cause of hospitalization.[11,12] Furthermore, joint injuries, such as anterior cruciate tear, that predispose to precocious joint failure have become epidemic in young athletes.[13,14] Although changes in the composition of synovial fluid have been described in patients with joint disease, there do not exist reliable biomarkers for early disease diagnosis or biomarkers that accurately depict response to therapy.[1,15−17] As a consequence, OA is often not diagnosed before irreversible damage has occurred.[15,18,19] Since synovial fluid is in direct contact with the joint tissues, it provides an attractive source of biomarkers candidates for monitoring joint health and for furthering the understanding of the disease mechanisms.[3,18] Previous studies of synovial fluid have mainly been focused on various diseased states. However, the protein concentration, content, and volume of synovial fluid is known to change dramatically during active joint diseases, and few studies have been focused on synovial fluid in healthy

Received: June 12, 2014

state or the likely origins of the synovial fluid proteins.[3,17] Balakrishnan et al.[20] compared the human synovial fluid proteome from OA and RA patients. Synovial fluid was immune-depleted of the most abundant proteins, followed by liquid chromatography−tandem mass spectrometry (LC−MS/MS) analysis, which led to the identification of 575 proteins, of which 135 demonstrated a greater than 3-fold abundance change between the two groups. In a similar study, Mateos et al.[1] identified a total of 136 different proteins using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) protein separation, followed by in-gel trypsin digestion and protein identification by LC−matrix-assisted laser desorption/ionization time-of-flight (MALDI TOF/TOF) MS. When assessing the total number of different proteins to be expected in synovial fluid, analyzing 2D-PAGE images seems to be a valid approach. Chen et al.[10] analyzed synovial fluid from patients with joint swelling using several precipitation strategies and detected 456 protein spots on 2D-PAGE images. Smith et al.[21] detected 1000 protein spots on 2D-PAGE images of synovial fluid from RA patients, which was estimated to represent 300 individual proteins. However, neither of the two latter studies identified the proteins within the protein spots. 2D-PAGE in-gel digestion strategies, while highly informative, are not suitable for high-throughput proteome analyses.[19] Therefore, we investigated the use of in-solution proteomics strategies with the aim to increase high-throughput protein identification and quantitation. In this normative proteomics study of synovial fluid from healthy porcine knee joints, we report results from applying our optimized method, using porcine synovial fluid. Domesticated or minipigs are a highly relevant anatomically large animal model organism to study several human diseases, including acute joint injury, OA, and other inflammatory disease.[22−26] The homology between the porcine genome and the human genome is conserved to a much greater extent than that between human and mouse or other rodents.[27] Furthermore, the porcine model has previously been used to study anterior cruciate ligament injury and surgical and tissue engineering approaches for healing.[28−30] We also report mRNA sequence data from healthy porcine synovium that we used to estimate the contribution to the synovial fluid proteome from this tissue, and we compare this to the relative contribution from plasma (e.g., proteins secreted into plasma by the liver or by blood cells) and articular cartilage.

EXPERIMENTAL PROCEDURES

Collection of Synovium and Synovial Fluid Samples

Six adolescent Yucatan minipigs (Coyote CCI, Douglas, MA), aged 12−15 months, were obtained for use in this study. All minipigs were housed and monitored by the Animal Resources at Boston Children's Hospital (ARCH) and handled according to approved Institutional Animal Care and Use Committee (IACUC) protocols. Minipigs were acclimated to the ARCH environment for a minimum of 3 days prior to experimental handling.

Nonbloody synovial fluid was obtained from the joints by aspiration using a 21 gauge needle. Samples were centrifuged at 3000*g* at room temperature for 10 min to pellet and remove cells and cellular debris. In some cases, 3 mL of sterile saline was injected into the knee joint to facilitate fluid extraction; after saline injection the knee was bent 10 times to ensure homogeneous fluid distribution and mixing. The saline/synovial fluid mix was then processed as above. Following centrifugation, the supernatants were stored at −80 °C. Furthermore, a human synovial fluid sample was obtained from a RA patient according to an approved IRB protocol (IRB-P00006443) to evaluate the integrity of the UniProt *Sus scrofa* protein database.

Euthanasia of the animals was induced by intramuscular injection of atropine (0.04 mg/kg), Telazol (4.4 mg/kg), and xylazine (2.2 mg/kg) and finalized by intravenous injection of Fatal Plus (86 mg/kg). At the time of euthanasia, synovia from the knee joints of the hind limbs were harvested. Care was taken to sample only the synovial membrane without any subintimal structures, such as fat or blood vessels. Each tissue specimen was snap frozen in liquid nitrogen and stored at −80 °C.

Protein Concentration

Total protein concentration for each sample (diluted 1:30 in water) was determined for normalization of sample material using a colorimetric (Bradford) protein assay kit (Bio-Rad, Hercules, CA) according to the manufacturer's instructions, with bovine serum albumin used as the standard.

SDS-PAGE

Thirty micrograms of total synovial fluid protein was prepared for sodium dodecyl sulfate (SDS)-PAGE in Laemmli sample buffer (Bio-Rad, Hercules, CA) according to the manufacturer's instructions. SeeBlue Plus2 pre-stained standard (Invitrogen, Carlsbad, CA) was used as the protein molecular weight standard. The sample was fractionated using NuPAGE 4−12% Bis-Tris minigels (Invitrogen) at 150 V for 65 min in MOPS SDS-running buffer (Invitrogen). The gel was stained using Coomassie blue, SimplyBlue SafeStain (Invitrogen), according to manufacturer's instructions.

Synovial Fluid Protein Digestion

Three trypsin digest protocols were evaluated:

(1). Filter-Aided Sample Preparation (FASP) Digestion. Performed using the FASP protein digestion kit (Protein Discovery, San Diego, CA) according to manufacturer's instructions using 30 kDa cutoff spin filters. Ninety micrograms of total synovial fluid protein was digested overnight at 37 °C with 2 μg of sequencing grade modified trypsin (Promega, Fitchburg, MA). To assess the need of glycan removal when working with synovial fluid, 500 U peptide-*N*4-(*N*-acetyl-beta-glucosaminyl)-asparagine amidase (PNGase F) (New England BioLabs, Ipswich, MA) was added to these samples prior to the trypsin digestion step, and samples were incubated overnight at 37 °C, after which the normal FASP protocol was continued. After trypsin digestion, the samples were desalted with TARGA C18 columns (Nest Group, Southborough, MA) and resuspended in 5% acetonitrile (ACN) and 5% formic acid (FA) prior to analysis.

(2). Urea In-Solution Digestion. Performed according to Gallien et al.[31] Ninety micrograms of total synovial fluid protein was diluted with 8 M urea in 100 mM ammonium bicarbonate to a final volume of 25 μL. The sample was reduced with dithiothreitol at a final concentration of 12 mM for 30 min at 37 °C and alkylated with iodoacetamide at a final concentration of 40 mM for 1 h at room temperature in the dark. The samples were diluted with 100 mM ammonium bicarbonate to a total volume of 100 μL, 2 μg of trypsin was added, and the sample was digested overnight at 37 °C. The samples were desalted

dx.doi.org/10.1021/pr500587x | J. Proteome Res. XXXX, XXX, XXX−XXX

with TARGA C18 columns (Nest Group) and resuspended in 5% ACN, 5% FA prior to analysis.

(3). In-Gel Digestion. Three gel-lanes, each loaded with 150 μg total synovial fluid protein, were divided into 10 sections each and subjected to standard in-gel tryptic digestion as previously described,[32−34] followed by analysis.

Human Plasma Protein Digestion

Human plasma was acquired as part of an ongoing method optimization study using a deidentified, discarded cord plasma sample and thus are not considered research of human subjects. One hundred micrograms of plasma protein was digested using the FASP protein digestion kit (Protein Discovery, San Diego, CA) with a modification to the recommended protocol. This modification involved the use of a 10 kDa MWCO filter instead of the stock 30 MWCO filters. A trypsin/LysC mix (Promega, Madison, WI) was added to the FASP filter at 1:25 ratio for digestion. The samples were then incubated at 37 °C overnight, and the resulting peptides were recovered as recommended by the manufacturer's protocol. The peptides were desalted using Oasis HLB columns (Waters, Milford, MA) and resuspended in 2% ACN in 0.1% FA prior to analysis.

LC−MS/MS Measurement and Proteomics Data Analysis

Two different high-resolution/high-accuracy mass spectrometer systems were used for the shotgun proteomic analysis: (1) For post-translational modification (PTM) analysis and method optimization, synovial fluid samples were analyzed on a TripleTOF 5600 (AB Sciex, Framingham, MA) connected online with a nanoflow UPLC and a NanoFlex system (Eksigent/AB Sciex). The samples were loaded onto a 15 cm reversed-phase C18 200 μm chip with 2 μL/min in 100% solvent A (0.1% FA). The samples were then separated using a 15 cm reversed-phase C18 75 μm chip and eluted with a linear gradient of 2% solvent B (0.1% FA in ACN), which was raised to 35% solvent B over 120 min (60 min for in-gel digested samples) at a constant flow rate of 500 nL/min. (2) The six FASP digested synovial fluid samples used to determine the synovial fluid protein list and abundances as well as the trypsin-digested human plasma samples were analyzed on a Q Exactive (Thermo Scientific, Waltham, MA) connected online to an EASY-nUPLC 1000 (Thermo Scientific). The samples were loaded onto a 10.5 cm reversed-phase C18 PicoChip with a flow rate of approximately 1 μL/min in 98% solvent A and 2% solvent B and were eluted with eluent B using a linear gradient that was raised to 35% over 120 min at a constant flow rate of 300 nL/min.

The AB Sciex.wiff data files were analyzed using ProteinPilot 4.5 (rev. 1656, Paragon Algorithm 4.5.0.0). To identify the most commonly single observed PTMs, data files were searched in thorough mode with a focus on biological modifications in ProteinPilot to include more than 300 different PTMs.

The .raw data files from synovial fluid and human plasma analyzed on the Q Exactive were searched using MaxQuant 1.4.1.2.[35] All standard settings were employed with carbamidomethyl (C) as a static modification and deamidation (NQR), oxidation (M), and protein N-terminal acetylation included as variable modifications. Label-free quantitation of all proteins was performed in MaxQuant based on integrated precursor intensities. Protein abundances are represented as protein intensity-based absolute quantitation values (iBAQ) and are reported for all proteins having at least two quantifiable unique peptides in at least three LC−MS runs.[36]

The human plasma sample was searched against the UniProt *Homo sapiens* reference proteome database with isoforms (downloaded 7/18/2014, containing 89 032 entries). The porcine synovial fluid data was searched against the UniProt *S. scrofa* reference proteome database (downloaded 11/09/2013, containing 26 070 entries). The human RA synovial fluid, used to evaluate the UniProt *S. scrofa* database, was searched against all reviewed *H. sapiens* UniProt proteins (downloaded 08/10/2013, containing 20 277 entries). All proteins and peptides are reported below a 1% false discovery rate (FDR) cutoff, and protein posterior error probability (PEP, equivalent to expectancy) was investigated to ensure only confident protein identifications.[35] For the PTM analysis, the search results were analyzed using ProteinPilot Descriptive Statistics Template, version 3.001, and for the protein abundance analysis, the iBAQ values were analyzed using Perseus, version 1.4.1.3, and IBM SPSS Statistics (version 21). Venn diagrams were created with BioVenn[37] and Venny.[38]

Assignment of Formerly Glycosylated Asparagine Residues

Four criteria were required to assign N-glycosylation sites: (I) a 1% FDR cutoff to all peptide spectral matches (PSMs); (II) all site assignments required the presence of a consensus site (CS) for N-glycosylation, i.e., NX(S/T), where X may be any amino acid except proline; (III) once CS status was established for all peptide assignments, an asparagine deamidation at the asparagine within the CS was required; and (IV), finally, all true site assignments were required to come from sample preparations that were treated with PNGase F. The FDR of site assignment was estimated by evaluation of the random rate of site assignment among control samples that were not treated with PNGase F. In this way, the rate of PSMs leading to the identification of a deglycosylated peptide may be compared.

RNA Extraction

Total RNA was extracted from frozen synovium tissue using the PureLink RNA mini kit (Ambion, Austin, TX), treated with PureLink DNase I (Life Technologies, Grand Island, NY) according to the manufacturer's protocol, and quantified. Briefly, frozen tissue samples were placed in tubes containing 5 metal lysing beads (Metal Bead Lysing Matrix, MP Biomedicals, Solon, OH) and 1 mL of TRIzol (Life Technologies). Tissue was homogenized using a FastPrep24 (MP Biomedicals, Solon, OH) at 6 m/s for 40 s. This was repeated two additional times, with samples briefly placed in liquid nitrogen between homogenization runs. Following total RNA extraction and quantitation via spectrophotometry using a Nanodrop 1000 (Thermo Scientific, Wilmington, DE), samples were submitted for quality analysis on a Bioanalyzer model 2100 (Agilent Technologies, Santa Clara, CA). Each RNA sample had an RIN > 7.20, indicating that they were of sufficient quality for sequencing.[39]

Library Preparation

A mRNA library for each tissue sample was prepared as previously described using the TruSeq RNA sample preparation kit, v2 (Illumina, San Diego, CA).[40] Briefly, mRNA was enriched from total RNA, chemically fragmented, reverse-transcribed using random hexamers, and ligated to barcoded adapters per the manufacturer's instructions. The cDNA fragments were amplified via PCR, and the libraries were washed with AMPure XP beads (Beckman Coulter Inc., Danvers, MA) to remove primer dimers. One microliter aliquots from each library were run on a 4−20% TBE gel (Life

Technologies, Grand Island, NY) for verification. Equal amounts of DNA from separately barcoded cDNA libraries were pooled (n = 12 per lane) and sent for 50 base pair paired-end sequencing on an HiSeq 2000 (Illumina).

RNA-Seq Data Analysis

Reads were mapped to the pig genome (Susscr3) using RUM.[41] Known genes were annotated with an R script based on data available from ENSEMBL. Expression level of each gene was quantified by the reads per kilobase of exon per million mapped reads (RPKM). The repeatability of the RNA-seq data was measured by comparing libraries generated from the right and left knees of the same animal and calculating the Pearson correlation coefficient (R^2) with respect to the genome-wide RPKM values. The calculated RPKM values were averaged across 12 samples for each gene, and these mean values were used in ranking the genes with respect to their expression levels.

In order to estimate the likely source of proteins detected in the synovial fluid, we additionally reviewed mouse liver,[42] mouse blood,[40] and human articular cartilage (authors' unpublished data) RNA-seq data. On the basis of homology information retrieved from the ENSEMBL database, we matched average RPKM values for each transcript with the proteins detected in the synovial fluid, and we identified the proteins for which all four mRNA expression values were available. We then determined the subset of proteins with a signal peptide. Among these proteins, we identified those with at least 2-fold higher mRNA levels in the synovium compared to all of the other individual tissues, and we then performed the same calculations for liver, cartilage, and blood.

RESULTS AND DISCUSSION

Optimizing Preparation and Digestion Methods for Synovial Fluid

Initially, we determined whether precipitating the synovial fluid proteins increased the diversity of the detectable protein bands by Coomassie-stained SDS-PAGE. However, similar to Chen et al.,[10] we found that precipitation appeared to reduce polypeptide diversity rather than increase it (data not shown); thus, we chose not to use precipitated samples for our subsequent analyses. Likewise, because albumin can function as a protein carrier for a wide range of proteins, we chose not to deplete albumin or other abundant proteins from our samples to avoid changing protein abundance profiles and removing important proteins that might bind the depletion targets.[1,10,43−45] Zhou et al.[46] reported that removing albumin led to the depletion of many proteins in human serum, including clinically useful biomarkers.

We compared urea in-solution digestion to FASP to find the most efficient digestion method without prior prefractionation for synovial fluid and used the robust and proven in-gel digestion as a reference. Urea in-solution digestion yielded the lowest number of 127 identified proteins. In contrast, FASP and in-gel-based digestion each yielded 227 protein identifications, indicating similar performance of the two digestion techniques. However, 170 of the 227 identified proteins were found in at least two of the three synovial fluid samples when using the FASP protocol in contrast to only 145 of the 227 proteins in the case of the in-gel digestion protocol (Figure 1). Furthermore, despite substantial overlap among the data sets, roughly 26% of proteins found in at least two of the three synovial fluid samples with FASP were missed by in-gel digestion, and 12% of the in-gel digestion proteins were missed

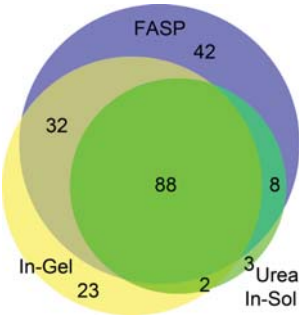


Figure 1. Digestion method evaluation. Number of proteins uniquely identified in at least two of the three synovial fluid samples, using FASP digestion, urea in-solution digestion, and in-gel digestion.

with FASP. Additionally, 17% of the proteins unique to the in-gel digestion were keratins, supporting the notion that in-gel digestion protocols can be prone to keratin contamination. Besides this difference, no other bias was found with regard to which groups of proteins were identified uniquely with the different digestion methods. This indicates that the FASP protocol gives more consistent data than the in-gel digestion protocol. It is possible that depletion strategies can increase the number of identified proteins further; however, this was not investigated in this study.

The UniProt *S. scrofa* reference proteome is not as well annotated as the human proteome. Therefore, in order to determine whether our proteomic method is as sensitive for detecting pig synovial fluid proteins as it would be for human synovial fluid proteins, we performed the same FASP digestion protocol on a synovial fluid sample from a RA patient. In the four technical replicates, we identified, on average, 173 different proteins in the human sample, which seems to be comparable to the 179 different proteins found, on average, in the individual porcine synovial fluid FASP LC−MS runs. The protein composition of synovial fluid is known to be altered during active joint diseases, so the protein overlap was not investigated.[3,17] Nonetheless, the similar number of identified proteins with the two databases indicates that the UniProt *S. scrofa* reference proteome is adequate for the analysis of the porcine synovial fluid.

Importance of Accounting for Post-Translational Modification in Protein Identification

Many PTMs occur in vivo or are introduced during sample preparation prior to MS analysis in vitro.[47−50] Therefore, to identify which PTMs should be considered for the database search of MS data from FASP-digested synovial fluid samples, we analyzed the identified peptides from this method in ProteinPilot Descriptive Statistics Template (Table 1). Several PTMs were identified, and all were likely artifacts from the sample preparation. Chloroacetamide can be used as an alternative alkylating agent to iodoacetamide to reduce the number of alkylation related artifacts.[47] However, due to the low number of detected peptides with these artificial modifications, this was not investigated further.

Synovial fluid is known to be rich in glycoproteins, which prompted us to evaluate the effect of removing N-linked glycans on the number of identifiable peptides and proteins.[2] We compared two synovial fluid samples using the FASP

dx.doi.org/10.1021/pr500587x | J. Proteome Res. XXXX, XXX, XXX−XXX

Table 1. Most Commonly Observed PTMs of the FASP-Digested Synovial Fluid Samples Based on the Most Frequent Single Features

rank	modification and position (N-terminal; C-terminal; acceptor amino acid)	modification mass (Da)	no. MS/MS events	modified sites of possible (%)
1	Carbamidomethyl (C)	57.0215	13 201	100.0
2	Deamidated (N)	0.9840	1024	8.7
3	Acetyl at N-term	42.0106	563	2.9
4	Carbamidomethyl at N-term	57.0215	535	2.7
5	Glu → pyro-Glu at N-term	−18.0106	522	24.1
6	Deamidated (Q)	0.9840	440	2.9
7	Carbamyl (K)	43.0058	434	2.0
8	Carbamidomethyl (K)	57.0215	420	1.9
9	Carbamyl at N-term	43.0058	416	2.1

protocol, with and without deglycosylation by PNGase F (±). As expected, PSMs resulting in identification of deglycosylated peptides considered to be previously N-glycosylated was increased ∼40-fold in the PNGase F-treated samples in comparison to that of the untreated samples (Table 2).

Table 2. Summary of Overall Numbers of PSMs in the Deglycosylation Experiments

sample	total no. of PSMs	PSM identifying deglycosylated peptides	deglycosylated peptide PSMs/all PSMs (%)
+ PNGase F 1	4827	91	1.9
+ PNGase F 2	4878	86	1.8
− PNGase F 1	4068	3	0.07
− PNGase F 2	4372	2	0.05

Twenty-eight and 34 unique deglycosylated peptides in the biological repeats were identified using the four criteria described in the Experimental Procedures section, compared to only two peptides in each of the untreated samples (Supporting Information Table 1). However, all deglycosylated peptides belonged to proteins that had already been identified from other peptides and thus resulted in no new protein identifications. Twenty-eight and 34 unique deglycosylated peptides are relatively low numbers for an analysis of a biological fluid known to be rich in glycoproteins, which may be due to a number of factors: First, the proteome of synovial fluid may not be particularly complex. This hypothesis is consistent with the proteomics data we observed in this study and with previously reported 2D-PAGE analyses.[1,10,17] Second, N-glycosylation, which is the only glycosylation amenable to PNGase F treatment protocols, may not be highly prevalent in synovial fluid. Synovial fluid contains significant amounts of glycosaminoglycans, possibly lessening the biologic necessity for N-glycosylation.[51,52] Third, the finite dynamic range of the MS analysis might cause abundant peptides to hinder the detection of less abundant ones. A study that exclusively focuses on O-glycosylation in synovial fluid and combines data from depleted and nondepleted samples will likely identify more extensive glycosylation than that reported here.

Identifying the Transcriptome of Healthy Pig Synovium

In order to identify synovial fluid proteins that are secreted by the synovium cells, i.e., transcribed and translated, we generated sequencing data of mRNA as active transcripts for pig synovium. We prepared and sequenced knee synovium from the right and left legs of 6 pigs. On average, we obtained 12.4 million sequencing reads per library, of which 90% mapped to the pig genome and 81% aligned uniquely. We then calculated RPKM values for individual genes and compared their expression level within each individual animal's left and right knee (Figure 2a). Intra-animal Pearson correlation coefficients exceeded 0.85, indicative of reproducible data (Figure 2d). The top 20 abundant protein-coding mRNAs that did not originate from the ribosome or mitochondria exhibited minimal variation in abundance, as they ranked in the top 1% of all sample-specific RNA-seq data sets that we analyzed (Figure 2c).

Focusing on transcripts encoding proteins containing signal peptides, which are most likely secreted or targeted to cell surfaces, we observed high expression of known connective tissue proteins such as decorin (DCN), fibronectin (FN1), collagen type III (COL3A1), and clusterin (CLU) (Figure 2b). Importantly, the principal lubricating protein in synovial fluid, lubricin (PRG4) made by type B synoviocytes, was also highly expressed. From these data, we produced a pig synovium transcriptome database (Supporting Information Table 2), which is composed of transcripts encoding known and predicted proteins.

Identifying the Proteome of Healthy Pig Synovial Fluid

We opted for the FASP digestion strategy for the detailed triplicate proteomic analysis of synovial fluid from 6 healthy pigs. This choice was based on the number of identified proteins, sample amount requirement, processing time, and instrument time. We identified 374 different proteins in total using the UniProt *S. scrofa* database and the synovium transcriptome database, and 42 proteins were identified solely in the transcriptome database. This seemed to be a large fraction assuming an adequate performance of the UniProt pig database, so all identified proteins were sorted by PEP (equivalent to expectancy[35]), with a high PEP value indicating a lower confidence in the identification (Figure 3a). Proteins unique to the transcriptome database displayed high PEP values, indicating that the number of identified proteins is too low for the global FDR calculation. To address this issue, all proteins with PEP > 1×10^{-3} were removed from further analysis, which came to 92 identified proteins, of which 36 were unique to the transcriptome database. Furthermore, we manually assessed the fragment spectra for proteins identified based on a single peptide and removed poor spectra with many unassigned high-intensity signals, which resulted in one additional protein being removed. After these two filtering steps, 267 proteins remained, of which 6 were unique to the transcriptome database (Figure 3b). Five of the 6 unique proteins were immunoglobulin lambda like proteins (UPID: IGLV-7 to IGLV-11), and the identified peptides originate from a variable region of the lambda chain, which is not annotated in the UniProt database. The remaining protein, ribonuclease 4 (UPID: I3LDZ2_PIG), was present in the UniProt database, but it had not been identified. We included several keratins (10, 14, 3, and 75) in the synovial fluid proteome because keratin-coding mRNA was detected in the articular cartilage. However, we cannot preclude some of the keratins being contaminants from sample preparation.

Of the identified 267 different proteins in synovial fluid (Supporting Information Table 2), 194 proteins (73%) were identified in all pigs, indicating similar protein expression patterns between the pigs and a high method robustness. The

dx.doi.org/10.1021/pr500587x | J. Proteome Res. XXXX, XXX, XXX−XXX

(a)

Library #	Number of Reads	Mapped (%)	Uniquely Mapped (%)
P1	11,040,002	92.3	83.4
P2	13,478,370	87.6	79.1
P3	11,536,658	89.9	81.1
P4	15,433,739	92.9	83.6
P5	11,443,699	87.6	79.0
P6	14,047,352	92.3	83.5
P7	13,243,709	84.2	76.0
P8	11,825,018	92.6	83.6
P9	11,046,003	89.1	79.9
P10	11,341,586	90.5	81.1
P11	13,169,629	92.2	83.8
P12	11,402,283	90.5	81.7
Average	12,417,337	90.1	81.3

(b)

Gene Symbol	Ensembl Symbol	Mean RPKM
EF1ALPHA	ENSSSCG00000004489	7138
DCN	ENSSSCG00000000915	6609
UBC	ENSSSCG00000029226	4898
CYTB	ENSSSCG00000018094	4684
FN1	ENSSSCG00000016174	4519
COL3A1	ENSSSCG00000016034	4141
CLU	ENSSSCG00000009668	4050
APOR	ENSSSCG00000015663	3910
uncharacterized gene	ENSSSCG00000000207	3501
MGP	ENSSSCG00000000606	3431
CST3	ENSSSCG00000029889	3268
PRG4	ENSSSCG00000029449	2940
FTH1	ENSSSCG00000014540	2796
TMSB10	ENSSSCG00000008245	2683
VIM	ENSSSCG00000011033	2547
FTL	ENSSSCG00000003153	2532
uncharacterized gene	ENSSSCG00000004579	2334
C1QC	ENSSSCG00000003525	2309
COL1A1	ENSSSCG00000017581	2297
S100A6	ENSSSCG00000021084	1933

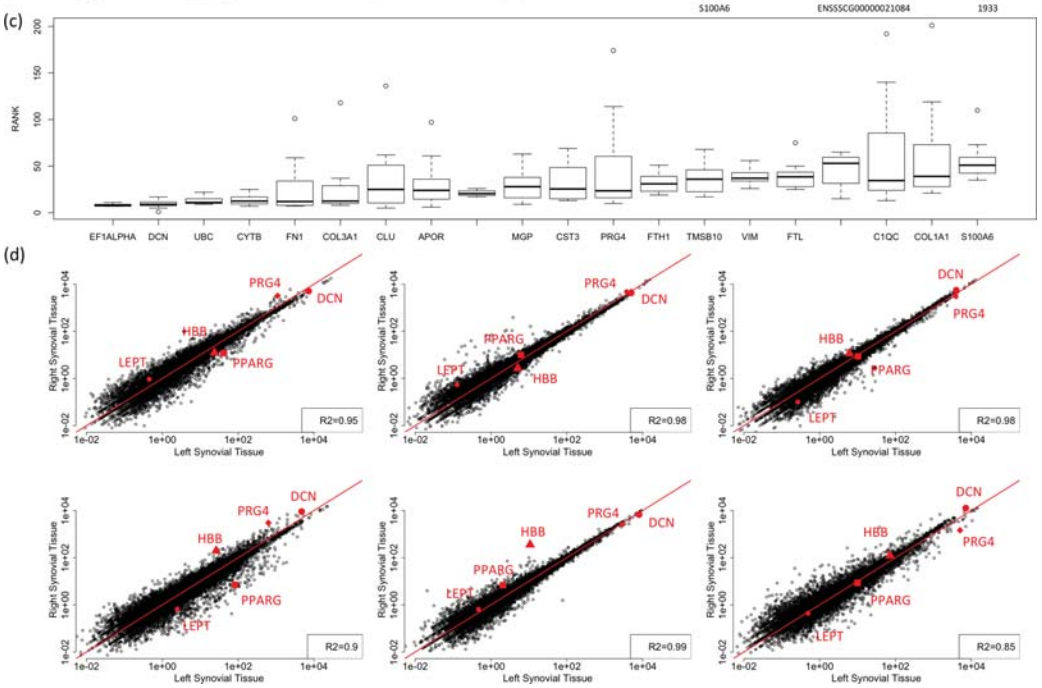


Figure 2. RNA-seq of pig synovium is repeatable and indicates high expression of several transcripts encoding secreted proteins. (a) Table indicating the total number of reads and percentages of successfully mapped reads. More than 90% of reads on average are mapped to the pig genome (Susscr3), and more than 80% of reads are mapped uniquely. (b) Top 20 protein coding genes (based on expression level measured with RPKM) that do not originate from the mitochondria or the ribosome. (c) Box plot indicating the variation in the expression-based rank order of genes listed in (b) in all 12 libraries analyzed. The highest variation is in the rank of PRG4; however, it remains within the top 1% of genes in all libraries. (d) Scatterplots indicating high similarity ($R^2 > 0.85$) between the synovial transcriptomes of paired left and right legs of all pigs. Data indicate RPKM, and each dot represents a single gene. Transcripts for the secreted proteins decorin (DCN) and lubricin (PRG4), red blood cell-derived beta hemoglobin (HBB), an adipocyte-derived transcription factor (PPARG), and secreted protein leptin (LEPT) are indicated with red colored symbols and closely follow the $y = x$ line (solid red line) in each plot. Importantly, transcripts for matrix-degrading enzymes and inflammatory cytokines are not abundant, clustering in the bottom left corner of the panels.

difference in identified proteins between the pigs likely originates from biological diversity, which is to be expected, and the finite dynamic range of the MS analysis.[53] To investigate the relative protein abundances in synovial fluid, we investigated individual synovial fluid protein abundance by calculating iBAQ values.[36] The protein abundance in synovial fluid, as estimated by the 203 iBAQ quantifiable proteins, spans 5 orders of magnitude (Figure 4). As expected, albumin was the most abundant protein, and no apparent bias could be found regarding the abundances of serum-derived proteins (RPKM < 5) compared to proteins expressed by the synovium in the joint (RPKM > 5) (see below).

Estimating the Relative Contributions of Plasma, Synovium, and Cartilage to the Synovial Fluid Proteome

Synovial fluid is a transudate of blood that has additional components added by the surrounding joint tissue including synovium, cartilage, and, in the knee, ligaments and menisci. To determine whether a likely tissue source for a synovial fluid protein can be predicted, we compared protein expression with

dx.doi.org/10.1021/pr500587x | J. Proteome Res. XXXX, XXX, XXX−XXX

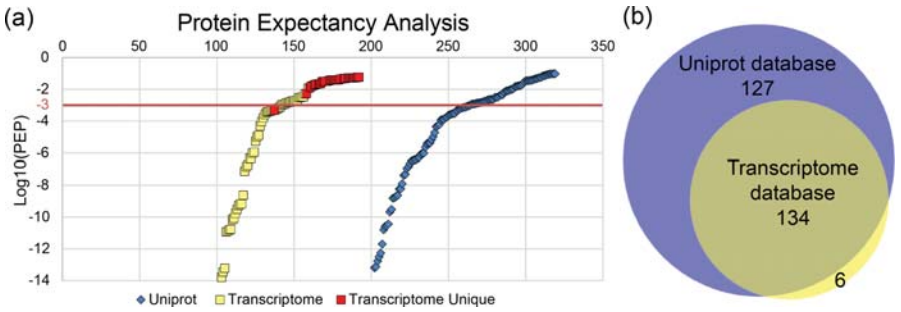


Figure 3. Analysis of identified synovial fluid proteins. (a) Identified proteins at 1% FDR sorted by PEP. Proteins unique to the transcriptome database (red squares) were, for the most part, identified with low confidence; hence, all proteins with PEP > 1×10^{-3} (red line) were removed from further analysis. (b) Synovial fluid proteins identified using the UniProt database and the transcriptome database fulfilling the filtering criteria. Of the 267 synovial fluid proteins, 140 had RPKM > 5 in the synovial mRNA libraries, indicating they are expressed by synovium.

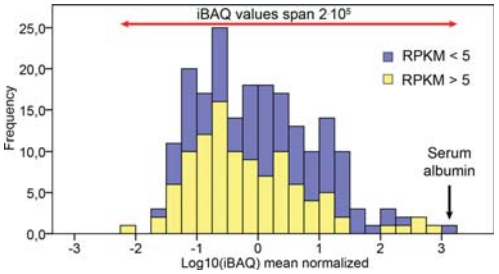


Figure 4. Protein abundance distribution in synovial fluid. Frequency histogram of the calculated iBAQ values for all proteomics quantifiable proteins. The relative proportion of proteins whose mRNA transcript abundances were less than or greater than 5 RPKM in the synovial membrane transcriptome are colored blue and yellow, respectively.

the pig synovium transcriptome, a human articular cartilage transcriptome we previously generated, and published mouse liver[42] and blood[40] transcriptomes (Supporting Information Table 2). As expected, albumin in synovial fluid solely derives from blood, as its transcript is highly expressed only in liver (Figure 5).

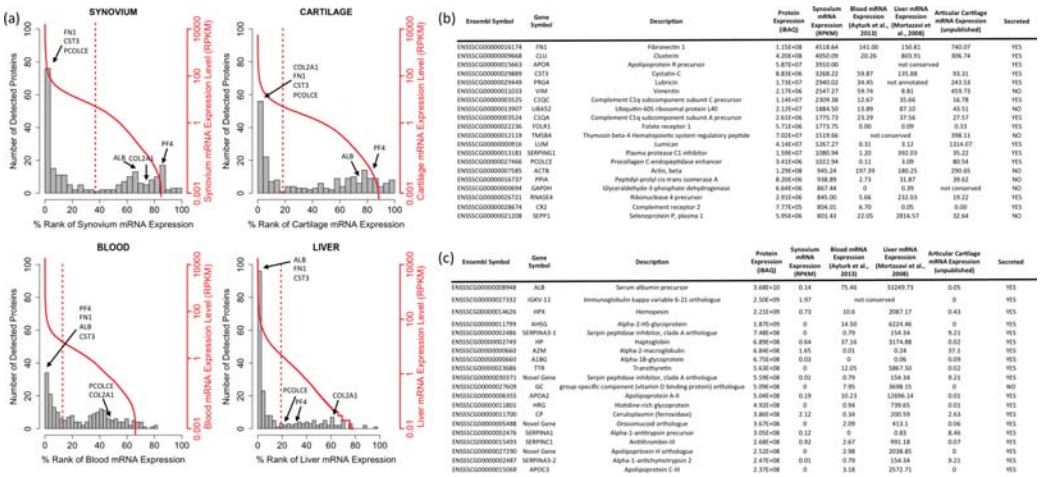
When we compared the proteins detected in the synovial fluid with the mRNA expression levels in four possible sources, we found that the majority of proteins corresponded to highly expressed transcripts regardless of the source and that liver and synovium likely made the greatest contribution (Figure 5a). Further analysis of our data revealed that 171 proteins detected in the synovial fluid had a signal peptide and that liver, blood, and articular cartilage mRNA expression data were available for 149 proteins (Supporting Information Table 2). Of these 149 common proteins in the synovial fluid proteome, 113 most likely derive from plasma transudate since their transcript abundance is 2-fold or higher in liver or blood than that in synovium or cartilage. In contrast, 37 synovial fluid proteins have transcripts whose expression in synovium is at least 2-fold higher than that in other tissues, while 11 proteins have at least 2-fold or higher transcript expression in articular cartilage. Several proteins serve as positive controls in validation of our estimations. For example, aggrecan, type II collagen, cartilage oligomeric matrix protein, and cartilage intermediate layer protein 2 are detected in the synovial fluid, and the corresponding transcripts are expressed at greater levels in articular cartilage than in other tissues. Similarly, proteins involved in post-translational modification of type I collagen (such as procollagen C-endopeptidase enhancer) and connective tissue markers (such as fibronectin 1 and clusterin) were expressed at higher levels in synovium than in other tissues. Thus, the putative origins of a substantial number of synovial proteome constituents can be predicted from the RNA sequencing data. For those 17 proteins whose mRNAs are expressed by multiple tissues, tissue-specific changes in mRNA abundance will need to be correlated with changes in synovial protein abundance to reliably determine their most likely tissue of origin.

We have additionally detected several proteins without a signal peptide, which are likely residual fragments that remained in the synovial fluid following apoptosis. Some examples of these are hemoglobin (HBA and HBB), which are expressed at extremely higher levels in blood than the remaining tissues, and beta-actin, a ubiquitously expressed protein whose mRNA was detected at high levels (RPKM > 180) in all four tissues. Also of interest are proteins not found in healthy pig synovial fluid and expressed at low levels in healthy synovium. Increased amounts of matrix metalloproteinases, including MMP13 and ADAMTS4, have been reported in synovial fluid from humans with OA, while proteins involved in the inflammatory/immune response, including complement activation, were found in patients with RA.[1] These proteins were neither detected in healthy pig synovial fluid nor were their transcripts abundant in healthy synovium. Therefore, in addition to monitoring changes in protein abundance for the healthy synovial fluid proteome, it will be of interest to compare proteomes and transcriptomes of healthy and diseased joints to identify pathways that may be integral to disease processes or mechanisms. Previous comparisons of synovial fluid obtained from patients with RA and OA identified 135 proteins at least 3-fold differentially abundant between the two groups.[20]

Comparison of Synovial Fluid and Plasma

Finally, to identify similarities between synovial fluid and blood plasma, from which many of the synovial fluid components were found to originate, we characterized the human plasma proteome (Supporting Information Table 3) and identified and quantified a total of 168 human plasma proteins. The detected iBAQ values of the human plasma proteins span 5 orders of

dx.doi.org/10.1021/pr500587x | J. Proteome Res. XXXX, XXX, XXX−XXX



(b)

Ensembl Symbol	Gene Symbol	Description	Protein Expression (iBAQ)	Synovium mRNA Expression (RPKM)	Blood mRNA Expression (Ayturk et al., 2013)	Liver mRNA Expression (Mortazavi et al., 2008)	Articular Cartilage mRNA Expression (unpublished)	Secreted
ENSSSCG00000016174	FN1	Fibronectin 1	1.15E+08	4518.64	141.00	150.81	740.07	YES
ENSSSCG00000009668	CLU	Clusterin	4.20E+08	4050.09	20.26	803.91	306.74	YES
ENSSSCG00000025663	APOA	Apolipoprotein A precursor	5.87E+07	3910.00		not conserved		YES
ENSSSCG00000029889	CST3	Cystatin-C	8.83E+06	3268.22	59.87	135.88	93.31	YES
ENSSSCG00000029449	PRG4	Lubricin	1.73E+07	2940.02	34.45	not annotated	243.53	YES
ENSSSCG00000011033	VIM	Vimentin	2.17E+06	2547.27	59.76	8.81	459.73	NO
ENSSSCG00000003525	C1QC	Complement C1q subcomponent subunit C precursor	1.14E+07	2309.38	12.67	35.66	16.78	YES
ENSSSCG00000013907	UBA52	Ubiquitin-60S ribosomal protein L40	2.12E+07	1884.50	13.89	87.10	43.51	NO
ENSSSCG00000003524	C1QA	Complement C1q subcomponent subunit A precursor	2.61E+06	1775.73	23.29	37.56	27.57	YES
ENSSSCG00000022236	FOLR1	Folate receptor 1	5.71E+06	1773.75	0.00	0.09	0.33	YES
ENSSSCG00000012118	TMSB4	Thymosin beta-4 Hematopoietic system regulatory peptide	7.02E+07	1519.66		not conserved	398.11	NO
ENSSSCG00000009916	LUM	Lumican	4.14E+07	1267.27	0.31	3.12	1314.07	YES
ENSSSCG00000013181	SERPING1	Plasma protease C1 inhibitor	1.59E+07	1080.94	1.20	392.03	35.22	YES
ENSSSCG00000027466	PCOLCE	Procollagen C-endopeptidase enhancer	3.41E+06	1022.94	0.11	6.09	80.54	YES
ENSSSCG00000007585	ACTB	Actin, beta	1.29E+08	945.24	197.39	180.25	290.65	NO
ENSSSCG00000016737	PPIA	Peptidyl-prolyl cis-trans isomerase A	8.20E+06	938.89	2.73	31.87	39.62	NO
ENSSSCG00000030694	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	6.64E+06	867.44	0	0.39	not conserved	NO
ENSSSCG00000026721	RNASE4	Ribonuclease 4 precursor	2.91E+06	845.00	5.66	232.03	19.22	YES
ENSSSCG00000028674	CR2	Complement receptor 2	7.77E+05	804.01	6.70	0.05	0.00	YES
ENSSSCG00000021208	SEPP1	Selenoprotein P, plasma 1	5.95E+06	801.43	22.05	2814.57	32.64	NO

(c)

Ensembl Symbol	Gene Symbol	Description	Protein Expression (iBAQ)	Synovium mRNA Expression (RPKM)	Blood mRNA Expression (Ayturk et al., 2013)	Liver mRNA Expression (Mortazavi et al., 2008)	Articular Cartilage mRNA Expression (unpublished)	Secreted
ENSSSCG00000008948	ALB	Serum albumin precursor	3.68E+10	0.14	75.46	33249.73	0.05	YES
ENSSSCG00000027332	IGKV-11	Immunoglobulin kappa variable 8-21 orthologue	2.50E+09	1.97	not conserved		0	YES
ENSSSCG00000014626	HPX	Hemopexin	2.21E+09	0.73	10.6	2087.17	0.43	YES
ENSSSCG00000011799	AHSG	Alpha-2-HS-glycoprotein	1.87E+09	0	14.50	6224.46	0	YES
ENSSSCG00000002486	SERPINA3 1	Serpin peptidase inhibitor, clade A orthologue	7.48E+08	0	0.79	154.34	9.21	YES
ENSSSCG00000002749	HP	Haptoglobin	6.89E+08	0.64	37.16	3174.88	0.02	YES
ENSSSCG00000000660	A2M	Alpha-2-macroglobulin	6.84E+08	1.65	0.01	0.24	37.1	YES
ENSSSCG00000000660	A1BG	Alpha-1B-glycoprotein	6.75E+08	0.03	0	0.06	0.09	YES
ENSSSCG00000023686	TTR	Transthyretin	5.63E+08	0	12.05	5867.50	0.02	YES
ENSSSCG00000030371	Novel Gene	Serpin peptidase inhibitor, clade A orthologue	5.59E+08	0.01	0.79	154.34	9.21	YES
ENSSSCG00000027609	GC	group-specific component (vitamin D binding protein) orthologue	5.09E+08	0	7.95	3698.15	0	NO
ENSSSCG00000006355	APOA2	Apolipoprotein A-II	5.04E+08	0.19	10.23	12696.14	0.01	YES
ENSSSCG00000011801	HRG	Histidine-rich glycoprotein	4.92E+08	0	0.94	739.65	0.01	YES
ENSSSCG00000011706	CP	Ceruloplasmin (ferroxidase)	3.86E+08	2.12	0.34	200.59	2.63	YES
ENSSSCG00000005488	Novel Gene	Orosomucoid orthologue	3.67E+08	0	2.09	413.1	0.06	YES
ENSSSCG00000002476	SERPINA1	Alpha-1-antitrypsin precursor	3.05E+08	0.12	0	0.83	8.46	YES
ENSSSCG00000015493	SERPINC1	Antithrombin-III	2.68E+08	0.92	2.67	991.18	0.07	YES
ENSSSCG00000027290	Novel Gene	Apolipoprotein H orthologue	2.52E+08	0	2.98	2038.85	0	YES
ENSSSCG00000002487	SERPINA3-2	Alpha-1-antichymotrypsin 2	2.47E+08	0.01	0.79	154.34	9.21	YES
ENSSSCG00000015069	APOC3	Apolipoprotein C-III	2.37E+08	0	3.18	2572.71	0	YES

Figure 5. mRNA expression levels associated with the proteins detected with MS. (a) Tissue-specific histograms indicating the number of detected proteins corresponding to each transcript. Genes are ranked based on mRNA expression along the *x* axis, and the secondary *y* axis (in red) indicates the expression levels in RPKM (in log scale). (b) Table indicating the top 20 mRNAs in synovium with detectable protein products in synovial fluid. Gene expression data derived from published and unpublished studies on tissues with potential influence on protein expression repertoire of serum and synovial fluid are also presented in the three columns on the right. Blood and liver mRNA data originate from mice, and articular cartilage mRNA data originate from humans. (c) Table indicating the top 20 proteins with no detectable mRNA expression in the synovium: note that the majority of mRNA species corresponding to these proteins are abundantly expressed by liver cells.

Table 3. Top Ten Most Abundant Proteins in Human Plasma and Porcine Synovial Fluid Ranked from the Most Abundant[a]

rank	human plasma	human UPID	porcine synovial fluid	human UPID
1	Serum albumin	P02768	Serum albumin	P02768
2	Ig gamma-1 chain C region	P01857	Hemoglobin subunit alpha	P69905
3	Ig kappa chain C region	P01834	Ig lambda chain C region	P0CG04
4	Apolipoprotein A-I	P02647	Hemoglobin subunit beta	P68871
5	Ig lambda-2 chain C regions	P0CG05	Serotransferrin	P02787
6	Ig gamma-2 chain C region	P01859	Immunoglobulin lambda variable 8-61	Q5NV62
7	Serotransferrin	P02787	Immunoglobulin kappa variable 6−21	P01834
8	Alpha-1-antitrypsin	P01009	Apolipoprotein A-I	P02647
9	Apolipoprotein A-II	P02652	Hemopexin	P02790
10	Alpha-2-macroglobulin	P01023	Alpha-2-HS-glycoprotein	P02765

[a]As determined by iBAQ values.

magnitude, similar to what was found for the synovial fluid, demonstrating the large degree of protein abundance variation encountered in the two body fluids. However, the actual range of protein abundances in plasma and synovial fluid is expected to be several orders of magnitude larger, due to the limited dynamic range of the MS analysis.[54] Probing the top 10 most abundant proteins in both fluids, the most abundant protein is serum albumin, as expected (Table 3). Other shared high-abundance proteins include serotransferrin, apolipoprotein A, and several immunoglobulins. Serotransferrin and apoliopoprotein A are binding and transporting proteins of iron and lipids, respectively.[55,56] The high abundance of the transporter proteins and immunoglobulins points to the shared functions of the two body fluids as a transport media of cellular products. The main difference between the high-abundance proteins in the plasma and synovial fluid proteomes is the high abundance of hemoglobin in the synovial fluid, which is detected with a lower abundance in plasma (rank 25). In both fluids, the protein is a likely contaminant from red blood cells prior to fluid centrifugation. Fibrinogens, involved in blood clotting, were identified with high abundance in plasma in contrast to that in synovial fluid, as expected.

Our findings demonstrate the high degree of similarity shared by the synovial fluid proteome and the plasma proteome as well as the serum-derived nature of many synovial fluid components. The expected high degree of similarity found between human plasma and porcine synovial fluid, furthermore, indicates that porcine model systems of synovial joints are suitable for studies focusing on human joint diseases.[57]

■ CONCLUSIONS

We evaluated different proteomic strategies for analyzing healthy synovial fluid. We utilized porcine synovial fluid because this animal is used to model human joint disease. In this study, we developed a FASP-based analysis pipeline that consumes only 0.1% of the starting sample material and 10% of the MS instrument time used by, e.g., Balakrishnan;[20] thus, our method lends itself to clinical proteomics studies on larger

dx.doi.org/10.1021/pr500587x | J. Proteome Res. XXXX, XXX, XXX−XXX

cohorts. Applying this fast and efficient workflow, we completed a normative proteomics study of synovial fluid from healthy knee joints and identified more than 250 proteins using very stringent identification criteria. Preprocessing with PNGase F to remove N-glycosylation was not found to be essential for the proteomic analysis of synovial fluid. Through an analysis of high-occurrence peptide PTMs, we found that proteomics projects using the FASP protocol for synovial fluid analysis, and possibly for other biological samples as well, should include peptide N-terminal acetylation, deamidation of asparginine and glutamine, and peptide N-terminal pyroglutamic acid in the data analysis for comprehensive peptide identification.

Going beyond the conventional LC−MS-based proteome mapping, we also classified and predicted the likely tissue origins for the majority of the detected proteins in healthy synovial fluid by cross-referencing the proteome with the RNA transcriptomes from synovium, cartilage, blood, and liver. Although many proteins derive from plasma transudate, as expected, an important subfraction appears to be solely expressed by synovium or cartilage. Changes in the abundance of these latter proteins in synovial fluid or in blood may be useful biomarkers of disease onset or progression. For synovial fluid proteins whose transcripts are expressed in multiple tissues, it will be necessary to correlate tissue-specific changes in mRNA expression with changing synovial fluid protein abundance in order to determine these proteins' principal tissues of origin. Finally, as many proteins were found to likely originate from plasma, we compared the identified porcine synovial fluid proteome to the proteome of human plasma. We found that the two body fluids share many similarities in terms of protein functions and localizations, underlining the detected plasma derived nature of many synovial fluid components.

Our methods and the resulting proteome and transcriptome data sets will be useful when comparing porcine synovial fluid in healthy and diseased states. Importantly, the high homology between pigs and humans should make these methods and data sets valuable for human studies.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Table 1 contains all identified glycosylated peptides. Table 2 contains all identified proteins, iBAQ and RPKM values (when available), sequence coverage, and imported transcriptome information from associated mouse tissues. Table 3 contains all identified human plasma proteins. Table 4 contains additional information for all proteins identified from a single peptide. This material is available free of charge via the Internet at http://pubs.acs.org. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[58−61] with the data set identifier PXD000935.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +1-617-919-2629. Fax: +1-617-730-0148. E-mail: hanno.steen@childrens.harvard.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This investigation was supported by National Institutes of Health under NIAMS AR054099 (M.M.M.) and AR050180 (M.L.W.) and by the Ruth L. Kirschstein National Research Service Award (F32 AR061186) (C.M.H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The overseas collaboration was supported by the Knud Hojgaards Foundation Denmark, the Lundbeck Foundation Denmark, the Oticon Foundation Denmark, and the Otto Monsteds Foundation Denmark. The authors would like to thank the ARCH staff, Dr. Arthur Nedder, Kathryn Mullen, Dana Bolgen, and Courtney White for their assistance and care in handling the minipigs, Saima Ahmed, Nino Esile, and Kevin Broadbelt for assistance in the proteomics laboratory, and Lau Sennels for help with the proteomics data analysis.

■ ABBREVIATIONS

2D-PAGE, two-dimensional polyacrylamide gel electrophoresis; ACN, acetonitrile; CLU, clusterin; COL3A1, collagen type III; CS, consensus site; DCN, decorin; FA, formic acid; FASP, filter-aided sample preparation; FDR, false discovery rate; FN1, fibronectin; HBA, alpha hemoglobin; HBB, beta hemoglobin; iBAQ, intensity-based absolute quantitation; LC, liquid chromatography; LEPT, secreted protein leptin; MALDI, matrix-assisted laser desorption/ionization; MS, mass spectrometry; OA, osteoarthritis; PEP, posterior error probability; PNGase F, peptide-*N*4-(*N*-acetyl-beta-glucosaminyl) asparagine amidase; PPARG, adipocyte-derived transcription factor; PRG4, lubricin; PSM, peptide spectral matches; PTM, post-translational modification; RA, rheumatoid arthritis; RPKM, reads per kilobase of exon per million mapped reads; SDS, sodium dodecyl sulfate; TOF, time-of-flight

■ REFERENCES

(1) Mateos, J.; Lourido, L.; Fernández-Puente, P.; Calamia, V.; Fernández-López, C.; Oreiro, N.; Ruiz-Romero, C.; Blanco, F. J. Differential protein profiling of synovial fluid from rheumatoid arthritis and osteoarthritis patients using LC−MALDI TOF/TOF. *J. Proteomics* **2012**, *75*, 2869−2878.

(2) Swann, D. A.; Hendren, R. B.; Radin, E. L.; Sotman, S. L.; Duda, E. A. The lubricating activity of synovial fluid glycoproteins. *Arthritis Rheum.* **1981**, *24*, 22−30.

(3) Hui, A. Y.; McCarty, W. J.; Masuda, K.; Firestein, G. S.; Sah, R. L. A systems biology approach to synovial joint lubrication in health, injury, and disease. *Wiley Interdiscip. Rev.: Syst. Biol. Med.* **2012**, *4*, 15−37.

(4) Schmid, K.; Macnair, M. B. Characterization of the proteins of certain postmortem human synovial fluids. *J. Clin. Invest.* **1958**, *37*, 708−718.

(5) Schmid, K.; Macnair, M. B. Characterization of the proteins of human synovial fluid in certain disease states. *J. Clin. Invest.* **1956**, *35*, 814−824.

(6) Ruiz-Romero, C.; Calamia, V.; Carreira, V.; Mateos, J.; Fernández, P.; Blanco, F. J. Strategies to optimize two-dimensional gel electrophoresis analysis of the human joint proteome. *Talanta* **2010**, *80*, 1552−1560.

(7) Ropes, M. W.; Rossmeisl, E. C.; Bauer, W. The origin and nature of normal human synovial fluid. *J. Clin. Invest.* **1940**, *19*, 795−799.

(8) Platt, D.; Pigman, W.; Holley, H. L.; Patton, F. M. An electrophoretic study of normal and post-mortem human and bovine synovial fluids. *Arch. Biochem. Biophys.* **1956**, *64*, 152−163.

(9) Levick, J. R. Permeability of rheumatoid and normal human synovium to specific plasma proteins. *Arthritis Rheum.* **1981**, *24*, 1550−1560.

(10) Chen, C.; Hsu, C. C.; Yeh, W. L.; Lin, H. C.; Hsieh, S. Y.; Lin, S. C.; Chen, T. T.; Chen, M.; Tang, S. Optimizing human synovial fluid preparation for two-dimensional gel electrophoresis. *Proteome Sci.* **2011**, *9*, 65.
(11) Murphy, L.; Helmick, C. G. The impact of osteoarthritis in the United States: a population-health perspective. *Am. J. Nurs.* **2012**, *112*, S13−S19.
(12) Lawrence, R. C.; Felson, D. T.; Helmick, C. G.; Arnold, L. M.; Choi, H.; Deyo, R. A.; Gabriel, S.; Hirsch, R.; Hochberg, M. C.; Hunder, G. G.; et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: part II. *Arthritis Rheum.* **2008**, *58*, 26−35.
(13) Dragoo, J. L.; Braun, H. J.; Durham, J. L.; Chen, M. R.; Harris, A. H. S. Incidence and risk factors for injuries to the anterior cruciate ligament in national collegiate athletic association football data from the 2004−2005 through 2008−2009 National Collegiate Athletic Association Injury Surveillance System. *Am. J. Sports Med.* **2012**, *40*, 990−995.
(14) Brown, T. D.; Johnston, R. C.; Saltzman, C. L.; Marsh, J. L.; Buckwalter, J. A. Posttraumatic osteoarthritis: a first estimate of incidence, prevalence, and burden of disease. *J. Orthop. Trauma* **2006**, *20*, 739−744.
(15) Mobasheri, A. Osteoarthritis year 2012 in review: biomarkers. *Osteoarthritis Cartilage* **2012**, 1−14.
(16) Ritter, S. Y.; Subbaiah, R.; Bebek, G.; Crish, J.; Scanzello, C. R.; Krastins, B.; Sarracino, D.; Lopez, M. F.; Crow, M. K.; Aigner, T.; et al. Proteomic analysis of synovial fluid from the osteoarthritic knee: comparison with transcriptome analyses of joint tissues. *Arthritis Rheum.* **2013**, *65*, 981−992.
(17) Gobezie, R.; Kho, A.; Krastins, B.; Sarracino, D. A.; Thornhill, T. S.; Chase, M.; Millett, P. J.; Lee, D. M. High abundance synovial fluid proteome: distinct profiles in health and osteoarthritis. *Arthritis Res. Ther.* **2007**, *9*, 36.
(18) Chiaradia, E.; Pepe, M.; Tartaglia, M.; Scoppetta, F.; D'Ambrosio, C.; Renzone, G.; Avellini, L.; Moriconi, F.; Gaiti, A.; Bertuglia, A.; et al. Gambling on putative biomarkers of osteoarthritis and osteochondrosis by equine synovial fluid proteomics. *J. Proteomics* **2012**, *75*, 4478−4493.
(19) Ruiz-Romero, C.; Blanco, F. J. Proteomics role in the search for improved diagnosis, prognosis and treatment of osteoarthritis. *Osteoarthritis Cartilage* **2010**, *18*, 500−509.
(20) Balakrishnan, L.; Bhattacharjee, M.; Ahmad, S.; Nirujogi, R. S.; Renuse, S.; Subbannayya, Y.; Marimuthu, A.; Srikanth, S. M.; Raju, R.; Dhillon, M. Differential proteomic analysis of synovial fluid from rheumatoid arthritis and osteoarthritis patients. *Clin. Proteomics* **2014**, *11*, 1.
(21) Smith, M. A.; Bains, S. K.; Betts, J. C.; Choy, E. H. S.; Zanders, E. D. Use of two-dimensional gel electrophoresis to measure changes in synovial fluid proteins from patients with rheumatoid arthritis treated with antibody to CD4. *Clin. Diagn. Lab. Immunol.* **2001**, *8*, 105−111.
(22) Yang, C.; Li, L.; Xue, Y.; Zhao, Z.; Zhao, T.; Jia, Y.; Rong, R.; Xu, M.; Nicholson, M. L.; Zhu, T. Innate immunity activation involved in unprotected porcine auto-transplant kidneys preserved by naked caspase-3 siRNA. *J. Transl. Med.* **2013**, *11*, 210.
(23) Maeyama, A.; Hoshino, Y.; Debandi, A.; Kato, Y.; Saeki, K.; Asai, S.; Goto, B.; Smolinski, P.; Fu, F. H. Evaluation of rotational instability in the anterior cruciate ligament deficient knee using triaxial accelerometer: a biomechanical model in porcine knees. *Knee Surg., Sports Traumatol., Arthrosc.* **2011**, *19*, 1233−1238.
(24) Wei, L.; Fleming, B. C.; Sun, X.; Teeple, E.; Wu, W.; Jay, G. D.; Elsaid, K. A.; Luo, J.; Machan, J. T.; Chen, Q. A Comparison of Differential Biomarkers of Osteoarthritis with and without Post-traumatic Injury in the Hartley Guinea Pig Model. *J. Orthop. Res.* **2010**, *28*, 900−906.
(25) Meurens, F.; Summerfield, A.; Nauwynck, H.; Saif, L.; Gerdts, V. The pig: a model for human infectious diseases. *Trends Microbiol.* **2012**, *20*, 50−57.
(26) De Almeida, A. M.; Bendixen, E. Pig proteomics: a review of a species in the crossroad between biomedical and food sciences. *J. Proteomics* **2012**, *75*, 4296−4314.
(27) Humphray, S. J.; Scott, C. E.; Clark, R.; Marron, B.; Bender, C.; Camm, N.; Davis, J.; Jenks, A.; Noon, A.; Patel, M.; et al. A high utility integrated map of the pig genome. *Genome Biol.* **2007**, *8*, R139.
(28) Murray, M. M.; Fleming, B. C. Use of a bioactive scaffold to stimulate ACL healing also minimizes post-traumatic osteoarthritis after surgery. *Am. J. Sports Med.* **2013**, *41*, 1762−1770.
(29) Vavken, P.; Fleming, B. C.; Mastrangelo, A. N.; Machan, J. T.; Murray, M. M. Biomechanical outcomes after bio-enhanced anterior cruciate ligament repair and anterior cruciate ligament reconstruction are equal in a porcine model. *Arthroscopy* **2012**, *28*, 672−680.
(30) Murray, M. M.; Magarian, E. M.; Harrison, S. L.; Mastrangelo, A. N.; Zurakowski, D.; Fleming, B. C. The effect of skeletal maturity on functional healing of the anterior cruciate ligament. *J. Bone Jt. Surg. Am.* **2010**, *92*, 2039−2049.
(31) Gallien, S.; Duriez, E.; Crone, C.; Kellmann, M.; Moehring, T.; Domon, B. Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol. Cell. Proteomics* **2012**, *11*, 1709−1723.
(32) Steen, H.; Küster, B.; Fernandez, M.; Pandey, A.; Mann, M. Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal. Chem.* **2001**, *73*, 1440−1448.
(33) Neubauer, G.; Mann, M. Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: potentials and limitations. *Anal. Chem.* **1999**, *71*, 235−242.
(34) Shevchenko, An.; Wilm, M.; Vorm, O.; Jensen, O. N.; Podtelejnikov, A. V.; Neubauer, G.; Shevchenko, Al.; Mortensen, P.; Mann, M. A strategy for identifying gel-separated proteins in sequence databases by MS alone. *Biochem. Soc. Trans.* **1996**, *24*, 893−896.
(35) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.
(36) Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. *Nature* **2011**, *473*, 337−342.
(37) Hulsen, T.; de Vlieg, J.; Alkema, W. BioVenn − a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **2008**, *9*, 488.
(38) J. C., Oliveros. VENNY. An interactive tool for comparing lists with Venn Diagrams http://bioinfogp.cnb.csic.es/tools/venny/.
(39) Schroeder, A.; Mueller, O.; Stocker, S.; Salowsky, R.; Leiber, M.; Gassmann, M.; Lightfoot, S.; Menzel, W.; Granzow, M.; Ragg, T. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **2006**, *7*, 3.
(40) Ayturk, U. M.; Jacobsen, C. M.; Christodoulou, D. C.; Gorham, J.; Seidman, J. G.; Seidman, C. E.; Robling, A. G.; Warman, M. L. An RNA-seq protocol to identify mRNA expression changes in mouse diaphyseal bone: applications in mice with bone property altering Lrp5 mutations. *J. Bone Miner. Res.* **2013**, *28*, 2081−2093.
(41) Grant, G. R.; Farkas, M. H.; Pizarro, A. D.; Lahens, N. F.; Schug, J.; Brunk, B. P.; Stoeckert, C. J.; Hogenesch, J. B.; Pierce, E. A. Comparative analysis of RNA-seq alignment algorithms and the RNA-seq unified mapper (RUM). *Bioinformatics* **2011**, *27*, 2518−2528.
(42) Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* **2008**, *5*, 621−628.
(43) Desrosiers, R. R.; Beaulieu, É.; Buchanan, M.; Béliveau, R. Proteomic analysis of human plasma proteins by two-dimensional gel electrophoresis and by antibody arrays following depletion of high-abundance proteins. *Cell Biochem. Biophys.* **2007**, *49*, 182−195.
(44) Schussler, G. C. The thyroxine-binding proteins. *Thyroid* **2000**, *10*, 141−149.
(45) Dea, M. K.; Hamilton-Wessler, M.; Ader, M.; Moore, D.; Schäffer, L.; Loftager, M.; Vølund, A.; Bergman, R. N. Albumin binding of acylated insulin (NN304) does not deter action to stimulate glucose uptake. *Diabetes* **2002**, *51*, 762−769.

(46) Zhou, M.; Lucas, D. A.; Chan, K. C.; Issaq, H. J.; Petricoin, E. F.; Liotta, L. A.; Veenstra, T. D.; Conrads, T. P. An investigation into the human serum "interactome". *Electrophoresis* **2004**, *25*, 1289−1298.
(47) Nielsen, M. L.; Vermeulen, M.; Bonaldi, T.; Cox, J.; Moroder, L.; Mann, M. Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat. Methods* **2008**, *5*, 459−460.
(48) Kollipara, L.; Zahedi, R. P. Protein carbamylation: in vivo modification or in vitro artefact? *Proteomics* **2013**, *13*, 941−944.
(49) Bennike, T.; Lauridsen, K. B.; Olesen, M. K.; Andersen, V.; Birkelund, S.; Stensballe, A. Optimizing the identification of citrullinated peptides by mass spectrometry: utilizing the inability of trypsin to cleave after citrullinated amino acids. *J. Proteomics Bioinform.* **2013**, *06*, 288−295.
(50) Bennike, T.; Birkelund, S.; Stensballe, A.; Andersen, V. Biomarkers in inflammatory bowel diseases: current status and proteomics identification strategies. *World J. Gastroenterol.* **2014**, *20*, 3231−3244.
(51) Bensouyad, A.; Hollander, A. P.; Dularay, B.; Bedwell, A. E.; Cooper, R. A.; Hutton, C. W.; Dieppe, P. A.; Elson, C. J. Concentrations of glycosaminoglycans in synovial fluids and their relation with immunological and inflammatory mediators in rheumatoid arthritis. *Ann. Rheum. Dis.* **1990**, *49*, 301−307.
(52) Nakano, T.; Aherne, F. X.; Thompson, J. R. Relative amounts of chondroitin sulfate and hyaluronic acid in synovial fluid from normal and osteochondrotic swine joints. *Can. J. Comp. Med.* **1984**, *48*, 434−436.
(53) Geiger, T.; Velic, A.; Macek, B.; Lundberg, E.; Kampf, C.; Nagaraj, N.; Uhlen, M.; Cox, J.; Mann, M. Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. *Mol. Cell. Proteomics* **2013**, *12*, 1709−1722.
(54) Hortin, G. L.; Sviridov, D. The dynamic range problem in the analysis of the plasma proteome. *J. Proteomics* **2010**, *73*, 629−636.
(55) Breslow, J. L.; Ross, D.; McPherson, J.; Williams, H.; Kurnit, D.; Nussbaum, A. L.; Karathanasis, S. K.; Zannis, V. I. Isolation and characterization of cDNA clones for human apolipoprotein A-I. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 6861−6865.
(56) Crichton, R. R.; Charloteaux-Wauters, M. Iron transport and storage. *Eur. J. Biochem.* **1987**, *164*, 485−506.
(57) Swindle, M. M.; Makin, A.; Herron, A. J.; Clubb, F. J.; Frazier, K. S. Swine as models in biomedical research and toxicology testing. *Vet. Pathol.* **2012**, *49*, 344−356.
(58) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32*, 223−226.
(59) Vizcaíno, J. A.; Côté, R. G.; Csordas, A.; Dianes, J. A.; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **2013**, *41*, D1063−1069.
(60) Wang, R.; Fabregat, A.; Ríos, D.; Ovelleiro, D.; Foster, J. M.; Côté, R. G.; Griss, J.; Csordas, A.; Perez-Riverol, Y.; Reisinger, F.; et al. PRIDE Inspector: a tool to visualize and validate MS proteomics data. *Nat. Biotechnol.* **2012**, *30*, 135−137.
(61) Côté, R. G.; Griss, J.; Dianes, J. A.; Wang, R.; Wright, J. C.; van den Toorn, H. W. P.; van Breukelen, B.; Heck, A. J. R.; Hulstaert, N.; Martens, L.; et al. The PRoteomics IDEntification (PRIDE) converter 2 framework: an improved suite of tools to facilitate data submission to the PRIDE database and the ProteomeXchange consortium. *Mol. Cell. Proteomics* **2012**, *11*, 1682−1689.

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-413-2

AALBORG UNIVERSITY PRESS